BACHELOR'S DEGREE THESIS

Degree in Mathematics

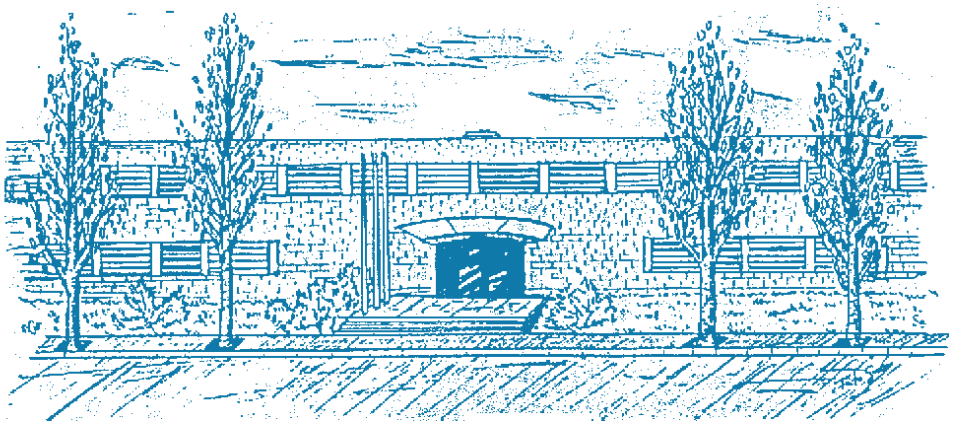
Title: **Modular symbols**

Author: **Gispert Sánchez, Francesc-Xavier**

Advisor: **Quer Bosor, Jordi**

Department: **Department of Mathematics**

Academic year: **2015/16**



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Modular symbols

Gispert Sánchez, Francesc-Xavier

under the direction of Quer Bosor, Jordi

A thesis submitted to the
Universitat Politècnica de Catalunya · BarcelonaTech
in partial fulfilment of the requirements for the
Bachelor's degree in Mathematics
and the
Bachelor's degree in Informatics Engineering

Catalonia
24th April 2016

Biblatex information:

```
@thesis{gispert2016modsymbols,
   author={Gispert Sánchez, Francesc},
   title={Modular symbols},
   date={2016-04-24},
   institution={Universitat Politècnica de Catalunya},
   type={Bachelor's thesis},
   pagetotal={123}
}
```

“Això és només àlgebra lineal de primer.”

Jordi Quer Bosor

Abstract

This thesis presents the classical theory of modular forms and modular symbols and explains the relations between these mathematical objects.

Modular forms are holomorphic functions defined on the complex upper half-plane which transform in a certain way under the action of some group of matrices. The orbit space of this action on the upper half-plane admits a structure of Riemann surface and so is called a modular curve. The spaces of modular forms are finite-dimensional complex vector spaces which can be identified with certain spaces of differential forms on the corresponding modular curve.

There is a very important family of operators acting on the space of modular forms, the Hecke operators. One of their main properties is that there exist bases of modular forms consisting of eigenvectors of most Hecke operators; these modular forms are known as eigenforms.

Finally, modular symbols can be thought of as formal symbols satisfying certain algebraic relations and which provide a simple way to represent elements of the first homology group of modular curves (regarded as compact surfaces). The pairing given by integration of a form along a path provides a duality between modular forms and modular symbols. Therefore, Hecke operators also act on the space of modular symbols. One can recover information about the modular forms from the action of Hecke operators on the modular symbols. In conclusion, modular symbols constitute an appropriate setting to perform computations with modular forms and Hecke operators.

Keywords: Hecke operators, modular curves, modular forms, modular symbols, number theory

MSC2010: 11F11, 11F67, 11F25

Preface

This thesis covers the classical theory of modular forms and introduces modular symbols with an emphasis on their computational aspects.

Many areas of mathematics come together in the theory of modular forms: complex analysis, algebraic topology, algebraic geometry and representation theory, to name just a few. Thus, modular forms arise naturally in many problems originating from a wide range of contexts in mathematics (and even in some branches of modern physics such as string theory). My interest, however, lies in the numerous applications of modular forms to number theory, a very active field of research in which this topic has gained much attention over the last decades.

Modular forms are analytic functions in the complex upper half-plane which transform in a certain way under the action of a group of matrices. Therefore, modular forms satisfy many symmetries which endow them with a very rich structure. In particular, modular forms have a Fourier series expansion. The Fourier coefficients of certain modular forms carry a large amount of arithmetic information. For instance, modular forms occur as generating functions of numbers of representations of integers by positive definite quadratic forms, special values of L–functions or invariants in algebraic number theory such as class numbers. But, without a doubt, one of the most celebrated arithmetic results involving modular forms is the modularity theorem, which states that every elliptic curve is associated with a modular form in some sense and illustrates a strong connection between modular forms and Galois representations. The proof by Wiles of this theorem for a large class of elliptic curves led to the conclusion of the proof of Fermat's last theorem after more than three centuries. All these are but a few examples of why modular forms play an essential role in modern number theory.

Despite their utmost importance, modular forms appear to be rather abstract objects and seem difficult to construct considering only their definition. Modular symbols are much more concrete objects which can be described algebraically. Thus, modular symbols provide a simple presentation of the space of modular forms with which one can perform all kinds of computations with ease. Also, modular symbols offer greater insight into the structure of modular forms, so they have been used to obtain some difficult results concerning modular

forms. Nevertheless, the theory of modular symbols is relatively unknown (in comparison with the theory of modular forms).

Professor Jordi Quer introduced me to this subject and proposed it as a topic for my Bachelor's degree thesis, arguing that it would be a great opportunity to learn at least some elementary aspects of a very active field of current research. As a matter of fact, the ultimate goal he had in mind was an open problem: finding an explicit way to express twists of modular forms by characters in the language of modular symbols in order to perform this kind of computations efficiently. Admittedly, this was too ambitious for a Bachelor's degree thesis, but the initial idea was to pose the problem and start thinking about it. In the end, I took a different approach and decided to study the subject of modular forms in greater detail, even at the cost of not getting that far in the theory of modular symbols.

In writing this thesis, I have made an effort to keep the prerequisites to a minimum. However, the knowledge which can be acquired in the Bachelor's degree (including the elective subjects) is assumed. In particular, a certain degree of understanding of complex analysis, abstract algebra, algebraic geometry and general and algebraic topology is required. In contrast, the theory of modular forms and modular symbols is explained from scratch.

Chapter 1 introduces the modular group and its action on the Poincaré upper half-plane in order to define modular forms. After giving the basic definitions, modular forms of level 1 are studied in detail: this case is so simple that modular symbols are not needed at all.

Chapter 2 describes the structure of Riemann surfaces of modular curves. This chapter involves a lot of geometry and topology but little arithmetic. Moreover, the proofs are quite technical (in fact, most of the references in the bibliography skip these proofs).

Chapter 3 explains Hecke operators from two different viewpoints: using modular points and using double cosets. In both approaches, Hecke operators are presented as very concrete objects by restricting the definitions to certain subgroups of matrices instead of explaining a much more general but abstract theory. This chapter plays a prominent role in the thesis because Hecke operators are the most important nexus between modular forms and modular symbols: a certain space of modular symbols constitutes a Hecke module which is dual to some space of modular forms.

In chapter 4, modular symbols are finally defined. The structure of the space

of modular symbols is completely determined with an algebraic presentation of the homology of the corresponding modular curve: this is the central result contained in this thesis. All this theory to characterise modular symbols and find the relations which they satisfy corresponds essentially to the first part of Manin's original paper [5] (although some explanations have been extended and reorganised here). Actually, most of the other references dealing with modular symbols do not include the proofs of the main facts (which are quite technical and involve, again, a lot of topology) and just cite this paper.

Finally, chapter 5 serves as a brief summary of the theory of modular symbols, greatly emphasising the computational aspects. Most of this chapter is devoted to the explanation of algorithms to compute the Fourier series of modular forms using modular symbols. These algorithms, along with the properties of modular symbols, are further illustrated with the detailed analysis of some examples computed using Sage [15].

I am indebted to Professor Jordi Quer for many things. First, for having introduced me to the fascinating subject of number theory in general and, more specifically, for the choice of the topic of this thesis. Second, for all the time he spent explaining things which I had not had the opportunity to learn before to me, even about topics which are not strictly related to the thesis. And last, but not least, for his careful reading of all this work and his valuable suggestions and comments, many of which have been incorporated in the final version of the thesis. Even so, needless to say, I alone am responsible for any deficiencies which may remain. These humble words cannot do justice to his indefatigable dedication, for which I am extremely grateful.

I would also like to express my gratitude to CFIS, its sponsors and all the people who make this project possible, for I have been able to study for two Bachelor's degrees at UPC for the last four years at essentially no cost (thanks to its excellence scholarships).

Francesc Gispert Sánchez

Barcelona, Catalonia
April 2016

Contents

Chapter 1

Modular forms (of level 1)

This chapter introduces the central objects of the thesis: modular forms. The basic definitions which appear in the classical theory of modular forms are introduced in general. Then, we study in more detail the situation for modular forms of level 1 (the easiest case in some sense) in order to provide the necessary motivation for the more technical constructions developed in further chapters.

The presentation of the material in this chapter follows closely Serre's excellent exposition in the last chapter of his book [11], complemented with the relevant parts of the books [3] by Koblitz, [4] by Lang and [2] by Diamond and Shurman. The motivation for certain definitions is based on Milne's notes [8].

1.1 The modular group

In this section, we explain the concepts which will lead to the definition of modular forms. Modular forms are a kind of analytic functions with a certain "invariance" condition (up to some factor). We focus first on their domain of definition and develop the language used in the proper definition of classical modular forms.

Definition 1.1. The *complex upper half-plane* or *Poincaré half-plane* is the set $\mathbb{H}$ of complex numbers with positive imaginary part:

$$\mathbb{H} = \{z \in \mathbb{C} : \mathfrak{Im}(z) > 0\}.$$

The *extended upper half-plane* is the union of $\mathbb{H}$ with the set of *cusps* $\mathbb{P}^1_{\mathbb{Q}} = \mathbb{Q} \cup \{\infty\}$, and we refer to it as $\mathbb{H}^*$.

We observe that $\mathbb{H}$ admits a natural structure of Riemann surface. Actually, it is one of the only three simply connected Riemann surfaces, up to biholomorphic isomorphism (the other two being the complex plane and the Riemann sphere).

Our interest at the moment, though, resides in the structure given by the action of certain multiplicative groups of matrices on these domains.

Several groups of matrices (sometimes regarded as groups of automorphisms of certain Riemann surfaces) are going to appear throughout this work, so we introduce some notation here. Let A be a commutative ring with identity (in this work, A will be one of $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$) and let $n \in \mathbb{N}$. The general linear group $\mathrm{GL}_n(\mathrm{A})$ is the group of $n \times n$ invertible matrices with entries in A with the operation of matrix multiplication. The special linear group $\mathrm{SL}_n(\mathrm{A})$ is the subgroup of $\mathrm{GL}_n(\mathrm{A})$ consisting of those matrices with determinant 1. We also consider the projective general linear group $\mathrm{PGL}_n(\mathrm{A})$ and the projective special linear group $\mathrm{PSL}_n(\mathrm{A})$: these groups are obtained as quotients of $\mathrm{GL}_n(\mathrm{A})$ and $\mathrm{SL}_n(\mathrm{A})$, respectively, by the subgroups consisting of the scalar matrices contained in their respective groups. That is, $\mathrm{PGL}_n(\mathrm{A}) = \mathrm{GL}_n(\mathrm{A})/\{\lambda \cdot 1 \in \mathrm{GL}_n(\mathrm{A}) : \lambda \in \mathrm{A}^\times\}$ and, similarly, $\mathrm{PSL}_n(\mathrm{A}) = \mathrm{SL}_n(\mathrm{A})/\{\lambda \cdot 1 \in \mathrm{SL}_n(\mathrm{A}) : \lambda \in \mathrm{A}^\times \text{ and } \lambda^n = 1\}$. (Notice that we use the symbol 1 to refer to both the identity element in A and to the identity matrix in $\mathrm{GL}_n(\mathrm{A})$: the context should make clear the intended meaning.) Moreover, if $\mathrm{A} \subseteq \mathbb{R}$, we write $\mathrm{GL}_n^+(\mathrm{A})$ for the subgroup of $\mathrm{GL}_n(\mathrm{A})$ consisting of those matrices with positive determinant. In this case as well, the orthogonal group $\mathrm{O}_n(\mathrm{A})$ is the subgroup of $\mathrm{GL}_n(\mathrm{A})$ consisting of orthogonal matrices (i.e., matrices whose transposes are equal to their inverses) and the special orthogonal group $\mathrm{SO}_n(\mathrm{A})$ is $\mathrm{SL}_n(\mathrm{A}) \cap \mathrm{O}_n(\mathrm{A})$.

Definition 1.2. The group $\mathrm{GL}_2(\mathbb{C})$ acts on $\mathbb{P}^1_{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ (which we identify with the Riemann sphere as a Riemann surface) by *linear fractional transformations* (also known as *Möbius transformations*) in the following way:

$$\begin{aligned} \mathrm{GL}_2(\mathbb{C}) \times \mathbb{P}^1_{\mathbb{C}} &\longrightarrow \mathbb{P}^1_{\mathbb{C}} \\ (\gamma, z) &\longmapsto \gamma(z) = \frac{az+b}{cz+d} \end{aligned}$$

where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

In the preceding definition, we adopt the convention that

$$\frac{a\infty + b}{c\infty + d} = \lim_{z \to \infty} \frac{az+b}{cz+d} = \frac{a}{c} \quad \text{and} \quad \frac{w}{0} = \infty \text{ for all } w \in \mathbb{P}^1_{\mathbb{C}}.$$

Now one checks easily that this indeed defines a left action. That is to say,

$$1(z) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) z = z$$

and

$$\gamma_1(\gamma_2(z)) = \frac{a_1\left(\frac{a_2z+b_2}{c_2z+d_2}\right)+b_1}{c_1\left(\frac{a_2z+b_2}{c_2z+d_2}\right)+d_1} = \frac{(a_1a_2+b_1c_2)z+(a_1b_2+b_1d_2)}{(c_1a_2+d_1c_2)z+(c_1b_2+d_1d_2)} = (\gamma_1\gamma_2)(z)$$

for all $z \in \mathbb{P}^1_{\mathbb{C}}$ and all $\gamma_1 = \left(\begin{smallmatrix} a_1 & b_1 \\ c_1 & d_1 \end{smallmatrix}\right), \gamma_2 = \left(\begin{smallmatrix} a_2 & b_2 \\ c_2 & d_2 \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{C})$.

We can consider the induced action of subgroups of $\mathrm{GL}_2(\mathbb{C})$ or quotients of $\mathrm{GL}_2(\mathbb{C})$ by normal subgroups which act trivially. Similarly, there is an action of such groups on stable subsets of $\mathbb{P}^1_{\mathbb{C}}$ too. In particular, we observe that a matrix $\gamma \in \mathrm{GL}_2(\mathbb{C})$ acts on $\mathbb{P}^1_{\mathbb{C}}$ in the same way as $\lambda\gamma$ for any $\lambda \in \mathbb{C}^\times$. Therefore, there is an induced action of $\mathrm{PGL}_2(\mathbb{C})$ on $\mathbb{P}^1_{\mathbb{C}}$ (these are precisely the automorphisms of the Riemann sphere).

We want to obtain an action on $\mathbb{H}$. In this case, we restrict the coefficients of the matrices to real numbers. That is, we consider the action of $\mathrm{GL}_2(\mathbb{R})$ on $\mathbb{P}^1_{\mathbb{C}}$ given by linear fractional transformations. As before, the action of a matrix is invariant under multiplication of the matrix by a non-zero scalar. In particular, for all $\gamma \in \mathrm{GL}_2(\mathbb{R})$ we can consider the matrix $\det(\gamma)^{-\frac{1}{2}} \cdot \gamma$ which acts in the same way as γ but has determinant ± 1. This means that we can focus solely on the action of matrices with determinant ± 1. Finally, a straight-forward computation yields the following result.

Lemma 1.3. *If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})$ and if $z = x + iy \in \mathbb{C}$ where $x, y \in \mathbb{R}$, then*

$$\gamma(z) = \frac{\left(ac|z|^2 + bd + (ad+bc)x\right) + i(ad-bc)y}{|cz+d|^2}$$

and, in particular,

$$\mathrm{Im}(\gamma(z)) = \det(\gamma) \cdot \frac{\mathrm{Im}(z)}{|cz+d|^2}.$$

The second assertion in lemma 1.3 tells us that $\mathbb{H}$ is stable under the action of $\mathrm{GL}_2^+(\mathbb{R})$ by linear fractional transformations and that, contrariwise, matrices with negative determinant map the upper half-plane to the lower half-plane. Thus, the (left) action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{H}$ by linear fractional transformations is well-defined.

Furthermore, the element $-1 \in \mathrm{SL}_2(\mathbb{R})$ acts trivially on $\mathbb{H}$, and no elements of $\mathrm{SL}_2(\mathbb{R})$ other than ± 1 do so. Indeed, if $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ acts trivially on $\mathbb{H}$, we have that $cz^2 + (d-a)z - b = 0$ for all $z \in \mathbb{H}$, which implies that $b = c = 0$ and $a = d = \pm 1$. Also, for every $z = x + iy \in \mathbb{H}$ (where $x, y \in \mathbb{R}$ and $y > 0$) there is a

matrix $\gamma = y^{-\frac{1}{2}} \cdot \left(\begin{smallmatrix} y & x \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ which maps i to z. We can then consider that it is the group $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm 1\}$ which operates, and this action is faithful and transitive (in fact, this is precisely the group of analytic automorphisms of $\mathbb{H}$). Hence, we usually identify the elements of $\mathrm{SL}_2(\mathbb{R})$ with their images in $\mathrm{PSL}_2(\mathbb{R})$ under the canonical projection. When we want to make this explicit, we will use a bar to denote the projection in $\mathrm{PSL}_2(\mathbb{R})$: we will write $\overline{\gamma}$ for the image of an element $\gamma \in \mathrm{SL}_2(\mathbb{R})$ in $\mathrm{PSL}_2(\mathbb{R})$ and $\overline{\Gamma}$ for the image of a subgroup $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$.

The classical theory of modular forms concerns itself with the action of certain subgroups of $\mathrm{SL}_2(\mathbb{R})$ and their images in $\mathrm{PSL}_2(\mathbb{R})$. Let us be more precise. $\mathrm{SL}_2(\mathbb{R})$ is a Lie group and, in particular, it is equipped with a topology (which coincides with the induced topology when we identify it with a subset of $\mathbb{R}^4$ in the obvious way). Therefore, $\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{PSL}_2(\mathbb{R})$ are topological groups. The interesting groups in our context are certain discrete subgroups of $\mathrm{SL}_2(\mathbb{R})$ which are called Fuchsian groups of the first kind. Nevertheless, we restrict our study even further to some subgroups arising in number theory (although there are others).

Definition 1.4. The group $\mathrm{SL}_2(\mathbb{Z})$ is called the *full modular group*.

$\mathrm{SL}_2(\mathbb{Z})$ is obviously a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$, and so are all its subgroups as a consequence.

In the next section, we will be interested in the behaviour of certain functions "at infinity" (we visualise ∞ as a point at the end of the imaginary axis and sometimes write $i\infty$ to make it explicit). But the elements of $\mathrm{SL}_2(\mathbb{Z})$ map ∞ to rational numbers. Even more is true: every rational number can be expressed as $\frac{a}{c}$ with $a, c \in \mathbb{Z}$ such that $(a, c) = 1$, and in this situation there exist $b, d \in \mathbb{Z}$ such that $ad - bc = 1$ (Bézout's identity); then, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \cdot \infty = \frac{a}{c}$. That is why the extended upper half-plane was defined to include the rational numbers $\mathbb{Q}$ as well as the point ∞. This means that $\mathrm{SL}_2(\mathbb{Z})$ acts on $\mathbb{H}^*$.

The subgroups studied in this work are subgroups of finite index of the modular group.

Definition 1.5. For any positive integer N, we define

$$\Gamma(\mathrm{N}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{\mathrm{N}} \right\}$$

and call it the *principal congruence subgroup* of level N. A *congruence subgroup* of $\mathrm{SL}_2(\mathbb{Z})$ is a subgroup Γ containing $\Gamma(\mathrm{N})$ for some N: the minimum such N is called the *level* of Γ.

Example 1.6. The most important families of congruence subgroups treated in this work are

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\}$$

and

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N} \right\}$$

for any positive integer N, which is the level of both subgroups. (Here, $*$ means any integer entry.)

$\Gamma(1) = \Gamma_0(1) = \Gamma_1(1) = \mathrm{SL}_2(\mathbb{Z})$ is the only congruence subgroup of level 1.

Proposition 1.7. *Let* N *be a positive integer. The principal congruence subgroup* $\Gamma(N)$ *is a normal subgroup of* $\mathrm{SL}_2(\mathbb{Z})$ *of finite index.*

Proof. By definition, $\Gamma(N)$ is the kernel of the morphism $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ obtained by reducing entries modulo N. Therefore, $\Gamma(N)$ is normal. Moreover, this morphism gives an exact sequence

$$1 \longrightarrow \Gamma(N) \hookrightarrow \mathrm{SL}_2(\mathbb{Z}) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$$

and the quotient group $\mathrm{SL}_2(\mathbb{Z})/\Gamma(N)$ must be isomorphic to some subgroup of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. We conclude that $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma(N)] \leq |\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})| < N^3 < \infty$. □

We obtain as an immediate corollary that all congruence subgroups have finite index in $\mathrm{SL}_2(\mathbb{Z})$.

As a matter of fact, the exact sequence in the proof of proposition 1.7 can be extended to a short exact sequence (the morphism $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is surjective). This provides a method of computing explicitly the index of $\Gamma(N)$ in $\mathrm{SL}_2(\mathbb{Z})$ (we can just count the number of elements in $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$), but we only need to know that it is finite.

Definition 1.8. Let Γ be a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$. A *fundamental domain* for the action of Γ on $\mathbb{H}$ is a closed subset D of $\mathbb{H}$ such that every orbit of Γ has an element in D and two points in D are in the same orbit only if they lie on the boundary ∂D. That is, every point $z \in \mathbb{H}$ is Γ–equivalent to a point in D, but no two distinct points z_1, z_2 in the interior $\mathring{\mathrm{D}}$ of D are Γ–equivalent.

Even if we do not require it in the definition, we usually want simply connected (or at least connected) fundamental domains. Thus, the fundamental domain

for a group Γ is "almost" a set of representatives of $\Gamma \backslash \mathbb{H}$ which, moreover, has a "reasonable" topological structure. Let us find fundamental domains for the congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$.

Theorem 1.9. *Let* $\mathrm{F} = \left\{ z \in \mathbb{H} : |z| \geq 1 \text{ and } |\mathfrak{Re}(z)| \leq \frac{1}{2} \right\}$ *and consider the matrices*

$$\mathrm{S} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathrm{T} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

(1) F *is a fundamental domain for* $\mathrm{SL}_2(\mathbb{Z})$. *Moreover, two distinct points* z *and* z' *of* F *are equivalent under* $\mathrm{SL}_2(\mathbb{Z})$ *if and only if*
   (i) $\mathfrak{Re}(z) = \pm\frac{1}{2}$ *and* $z = z' \pm 1$, *in which case* $z = \mathrm{T}(z')$ *or* $z' = \mathrm{T}(z)$, *or*
   (ii) $|z| = 1$ *and* $z' = -\frac{1}{z} = \mathrm{S}(z)$.

(2) *Let* $z \in \mathrm{F}$. *The stabiliser of* z *is* $\{\pm 1\}$ *except in the following three cases:*
   (i) $z = i$, *with* $\mathrm{SL}_2(\mathbb{Z})_i = \langle \mathrm{S} \rangle$, *so* $\mathrm{PSL}_2(\mathbb{Z})_i$ *has order* 2;
   (ii) $z = \rho = e^{\pi i/3}$, *with* $\mathrm{SL}_2(\mathbb{Z})_\rho = \langle \mathrm{TS} \rangle$, *so* $\mathrm{PSL}_2(\mathbb{Z})_\rho$ *has order* 3;
   (iii) $z = \rho^2 = e^{2\pi i/3}$, *with* $\mathrm{SL}_2(\mathbb{Z})_{\rho^2} = \langle \mathrm{ST} \rangle$, *so* $\mathrm{PSL}_2(\mathbb{Z})_{\rho^2}$ *has order* 3.

(3) $\mathrm{SL}_2(\mathbb{Z})$ *is generated by* S *and* T.

Proof. Let Γ be the subgroup of $\mathrm{SL}_2(\mathbb{Z})$ generated by S and T, and let $z \in \mathbb{H}$. We shall show that there exists some $\gamma \in \Gamma$ such that $\gamma(z) \in \mathrm{F}$. Recall that, if $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, then

$$\mathfrak{Im}(\alpha(z)) = \frac{\mathfrak{Im}(z)}{|cz + d|^2}.$$

Since c and d are both integers, the number of pairs (c, d) such that $|cz + d|$ is less than a given number is finite. Consequently, there exists some $\alpha \in \Gamma$ such that $\mathfrak{Im}(\alpha(z))$ is maximum among elements in the orbit Γz. Choose now some integer n such that $z' = \mathrm{T}^n(\alpha(z))$ satisfies that $-\frac{1}{2} \leq \mathfrak{Re}(z') \leq \frac{1}{2}$. I claim that z' belongs to F. Indeed, if that were not the case, we would have $|z'| < 1$ and

$$\mathfrak{Im}(\mathrm{S}(z')) = \mathfrak{Im}\left(\frac{-1}{z'}\right) = \frac{\mathfrak{Im}(z')}{|z'|^2} > \mathfrak{Im}(z') = \mathfrak{Im}(\alpha(z)),$$

which contradicts our choice of α. Therefore, the element $\gamma = \mathrm{T}^n \cdot \alpha$ has the desired property.

Now consider $z \in \mathrm{F}$ and $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that $\alpha(z) \in \mathrm{F}$ as well. Assume that $\mathfrak{Im}(\alpha(z)) \geq \mathfrak{Im}(z)$ (up to replacing (z, α) by $(\alpha(z), \alpha^{-1})$). This means that

$$|cz + d|^2 = (c\,\mathfrak{Re}(z) + d)^2 + (c\,\mathfrak{Im}(z))^2 \leq 1$$

and this is impossible if $|c| \geq 2$ (because $\mathfrak{Im}(z) \geq \frac{3}{4}$). This leaves only the cases $c = 0, 1, -1$.

- If $c = 0$, then $d = \pm 1$ and $\alpha = \left(\begin{smallmatrix} \pm 1 & b \\ 0 & \pm 1 \end{smallmatrix}\right)$, so $\alpha(z) = z \pm b$. Since both z and $\alpha(z)$ are in F, this implies that either $b = 0$ (and $\alpha = \pm 1$) or $b = \pm 1$. In this last case, one of the numbers $\mathfrak{Re}(z)$ and $\mathfrak{Re}(\alpha(z))$ must be equal to $-\frac{1}{2}$ and the other to $\frac{1}{2}$.
- If $c = 1$, the fact that $|z + d| \leq 1$ implies that $d = 0$ unless $z = \rho$, in which case $d = 0$ or -1, or $z = \rho^2$, in which case $d = 0$ or 1. If $d = 0$, we have that $|z| \leq 1$ and, therefore, $|z| = 1$; on the other hand, $\alpha = \left(\begin{smallmatrix} a & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $\alpha(z) = a - \frac{1}{z}$. In this situation, the first part of the discussion proves that $a = 0$ unless $\mathfrak{Re}(z) = \pm\frac{1}{2}$, which is to say, $z = \rho$ (and $a = 0$ or 1) or $z = \rho^2$ (and $a = 0$ or -1). If $d = -1$ and $z = \rho$, then $-a - b = 1$ and $\alpha(z) = \frac{a\rho+b}{\rho-1} = a - \frac{1}{\rho-1} = a + \rho$; this is only possible for $a = 0$ or -1. Similarly, if $d = 1$ and $z = \rho^2$, then $a - b = 1$ and $\alpha(z) = \frac{a\rho^2+b}{\rho^2+1} = a - \frac{1}{\rho^2+1} = a + \rho^2$; this is only possible for $a = 0$ or 1.
- Finally, the case $c = -1$ is reduced to the case $c = 1$ by changing the signs of a, b, c and d (this does not change $\alpha(z)$).

This completes the proof of (1) and (2) of the theorem.

It remains to prove that $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Let $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ and choose a point z_0 interior to F (for instance, $z_0 = 2i$). Consider $z = \alpha(z_0)$. We proved that there exists some $\gamma \in \Gamma$ such that $\gamma(z) \in \mathrm{F}$. Now z_0 and $\gamma(z) = (\gamma \cdot \alpha)(z_0)$ are equivalent under the action of $\mathrm{SL}_2(\mathbb{Z})$ and $z_0 \in \mathring{\mathrm{F}}$, so (1) tells us that $z_0 = (\gamma \cdot \alpha)(z_0)$ and $\gamma \cdot \alpha \in \mathrm{SL}_2(\mathbb{Z})_{z_0} = \{\pm 1\}$ by (2). That is, α and γ^{-1} are equal in $\mathrm{PSL}_2(\mathbb{Z})$. □

In fact, one can show that $\langle \overline{\mathrm{S}}, \overline{\mathrm{T}} \mid (\overline{\mathrm{S}})^2, (\overline{\mathrm{ST}})^3 \rangle$ is a presentation of $\mathrm{PSL}_2(\mathbb{Z})$. That is to say, $\mathrm{PSL}_2(\mathbb{Z})$ is the free product of $\langle \overline{\mathrm{S}} \rangle$ (cyclic of order 2) and $\langle \overline{\mathrm{ST}} \rangle$ (cyclic of order 3).

Figure 1.1 shows F and its transforms by the elements 1, T, TS, $\mathrm{ST}^{-1}\mathrm{S}$, S, ST, STS, $\mathrm{T}^{-1}\mathrm{S}$ and T^{-1} of $\mathrm{PSL}_2(\mathbb{Z})$. This kind of pictures become a useful tool for computing explicitly fundamental domains for congruence subgroups. The following result tells us how.

Proposition 1.10. *Let $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ and let F be the fundamental domain for Γ described in theorem 1.9. Let Γ' be a congruence subgroup of Γ and choose a set of representatives $\alpha_1, \ldots, \alpha_n$ of the left cosets of $\overline{\Gamma'}$ in $\overline{\Gamma}$, so that*

$$\overline{\Gamma} = \bigsqcup_{j=1}^{n} \overline{\alpha_j}\,\overline{\Gamma'} = \bigsqcup_{j=1}^{n} \overline{\Gamma'}\,\overline{\alpha_j^{-1}}.$$

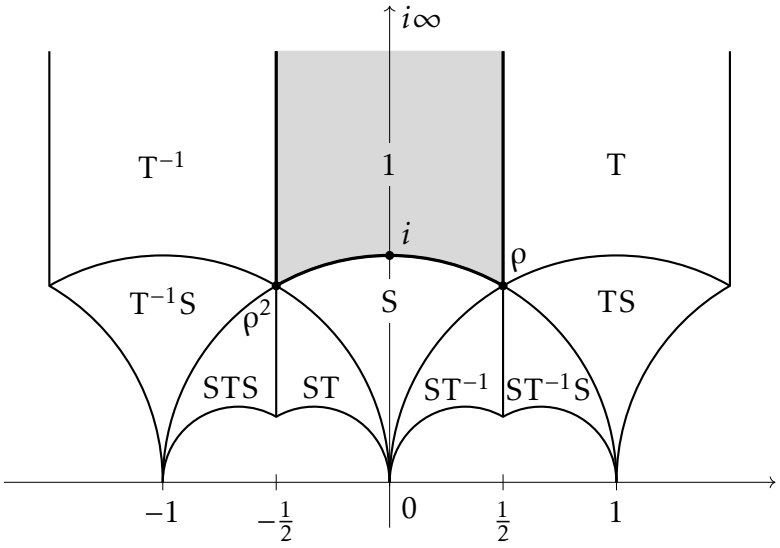


Figure 1.1: The fundamental domain F for $\mathrm{SL}_2(\mathbb{Z})$ described in theorem 1.9 and some of its transforms.

(Recall that a bar denotes the image in $\mathrm{PSL}_2(\mathbb{R})$*.) Then*

$$\mathrm{F}' = \bigcup_{j=1}^{n} \alpha_j^{-1}(\mathrm{F})$$

is a fundamental domain for Γ' *(possibly non-connected).*

Proof. Let $z \in \mathbb{H}$. Since F is a fundamental domain for Γ, there exist $z' \in \mathrm{F}$ and $\gamma \in \Gamma$ such that $z' = \gamma(z)$, and we can write $\gamma = \pm\alpha_j \cdot \gamma'$ for some $\gamma' \in \Gamma'$ and some j. Therefore, $\gamma'(z) \in \alpha_j^{-1}(\mathrm{F}) \subseteq \mathrm{F}'$.

If $z_1 = \gamma'(z_2)$ for some $z_1 \in \alpha_j^{-1}(\mathrm{F}) \subseteq \mathrm{F}'$ and $z_2 \in \alpha_k^{-1}(\mathrm{F}) \subseteq \mathrm{F}'$ and some $\gamma' \in \Gamma'$, then $\alpha_j(z_1) = \left(\alpha_j \cdot \gamma' \cdot \alpha_k^{-1}\right)(\alpha_k(z_2))$ and $\alpha_j(z_1), \alpha_k(z_2) \in \mathrm{F}$. Since F is a fundamental domain for Γ, either $\alpha_j \cdot \gamma' \cdot \alpha_k^{-1} = \pm 1$ (in which case $j = k$, $\gamma' = \pm 1$ and $z_1 = z_2$) or $\alpha_j(z_1), \alpha_k(z_2) \in \partial\mathrm{F}$. In the latter case, we have that $z_1 \in \partial\left(\alpha_j^{-1}\mathrm{F}\right)$ and $z_2 \in \partial\left(\alpha_k^{-1}\mathrm{F}\right)$ because the elements of $\mathrm{SL}_2(\mathbb{R})$ are diffeomorphisms of $\mathbb{H}$. We must prove that both z_1 and z_2 are in the boundary of F′ except if $z_1 = z_2$. To this aim, suppose that z_2 is interior to F′. Then there is an open neighbourhood U of z_2 in F′ and, as a consequence, $\gamma'(\mathrm{U})$ is an open neighbourhood of z_1 which might not be contained in F′. But at least we know that $\gamma'(\mathrm{U}) \cap \alpha_k^{-1}(\mathring{\mathrm{F}}) \neq \emptyset$. Therefore, we can choose $z_0 \in \mathring{\mathrm{F}}$ such that $\alpha_k^{-1}(z_0) \in \gamma'(\mathrm{U})$, so $\alpha_k^{-1}(z_0) = \gamma'\left(\alpha_l^{-1}(z_0')\right)$ for some $z_0' \in \mathrm{F}$ and some l (here $\alpha_l^{-1}(z_0') \in \mathrm{U} \subseteq \mathrm{F}'$). Again, since F is a fundamental domain, $z_0 = z_0'$, $\alpha_k^{-1} \cdot \gamma' \cdot \alpha_l^{-1} = -1$, $k = l$ and $\gamma' = \pm 1$. In particular, $z_1 = z_2$. □

The result of proposition 1.10 can be improved: one can choose $\alpha_1, \ldots, \alpha_n$ so that F′ is also connected. This gives an explicit way to compute a fundamental domain F′ for Γ′, as long as we can compute a set of representatives for $\mathrm{SL}_2(\mathbb{Z})/\Gamma'$. Fundamental domains are useful for visualising the geometric structure of the quotient $\Gamma' \backslash \mathbb{H}$. This is further formalised in chapter 2, where this set of orbits is given the structure of a Riemann surface.

1.2 Modular forms

Throughout this section, Γ will be a congruence subgroup of level N and k will be an integer. After introducing the action of Γ on $\mathbb{H}^*$ and having obtained a special set of representatives of the set of orbits $\Gamma \backslash \mathbb{H}$, the next natural step is to study functions which are invariant under the action of Γ. However, it is difficult to construct them directly. It is easier to construct functions which transform in a certain way under the action of Γ. Then, the quotient of two functions which transform in the same way will be invariant under the action of Γ.

This situation is analogous to how one might proceed in order to define functions on projective spaces. For example, let K be an infinite field and consider the projective line $\mathbb{P}^1_{\mathrm{K}} = ((\mathrm{K} \times \mathrm{K}) \setminus \{0\})/\mathrm{K}^\times$. Let $\mathrm{K}(\mathrm{X}, \mathrm{Y})$ be the field of fractions of $\mathrm{K}[\mathrm{X}, \mathrm{Y}]$. An element $f \in \mathrm{K}(\mathrm{X}, \mathrm{Y})$ defines a function $(a, b) \mapsto f(a, b)$ on some subset of $\mathrm{K} \times \mathrm{K}$ (where the denominator does not vanish), and this function passes to the quotient $\mathbb{P}^1_{\mathrm{K}}$ if and only if $f(\lambda \mathrm{X}, \lambda \mathrm{Y}) = f(\mathrm{X}, \mathrm{Y})$ for all $\lambda \in \mathrm{K}^\times$. We can obtain one such function if we consider first two homogeneous polynomials $g, h \in \mathrm{K}[\mathrm{X}, \mathrm{Y}]$ of the same degree d, so that $g(\lambda \mathrm{X}, \lambda \mathrm{Y}) = \lambda^d g(\mathrm{X}, \mathrm{Y})$ and $h(\lambda \mathrm{X}, \lambda \mathrm{Y}) = \lambda^d h(\mathrm{X}, \mathrm{Y})$. In this case, $f = g/h$ satisfies the desired condition.

The previous discussion motivates the definitions in this section.

Definition 1.11. We define a right action of weight k of $\mathrm{GL}_2^+(\mathbb{Q})$ on functions $f : \mathbb{H}^* \to \mathbb{P}^1_{\mathbb{C}}$ in the following way:

$$f|_k^{[\gamma]}(z) = \det(\gamma)^{\frac{k}{2}} j(\gamma, z)^{-k} f(\gamma(z))$$

where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$ and $j(\gamma, z) = cz + d$ is called the *automorphy factor*.

This indeed defines a right action: for all $\gamma_1, \gamma_2 \in \Gamma$,

$$j(\gamma_1 \gamma_2, z) = j(\gamma_1, \gamma_2(z)) j(\gamma_2, z)$$

and so

$$\left(f|_k^{[\gamma_1]}\right)\Big|_k^{[\gamma_2]}(z) = \det(\gamma_1)^{\frac{k}{2}}\det(\gamma_2)^{\frac{k}{2}} j(\gamma_1,\gamma_2(z))^{-k} j(\gamma_2,z)^{-k} f(\gamma_1(\gamma_2(z))) \\ = \det(\gamma_1\gamma_2)^{\frac{k}{2}} j(\gamma_1\gamma_2,z)^{-k} f((\gamma_1\gamma_2)(z)) = f|_k^{[\gamma_1\gamma_2]}(z).$$

Definition 1.12. We say that a function $f : \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ is *weakly modular* for Γ of *weight* k if $f|_k^{[\gamma]} = f$ for all $\gamma \in \Gamma$.

The functions we will be studying should have some desirable properties. Hence, we will impose two kinds of conditions: conditions in $\mathbb{H}$ and conditions at the cusps. Let us see what that means.

Observe that, since $\Gamma(N) \subseteq \Gamma$, there exists a positive integer h such that $T^h = \left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right) \in \{\pm 1\} \cdot \Gamma$ (namely, $h = N$). We can choose the minimum such h, which is called the width of the cusp ∞. Therefore, a meromorphic weakly modular function $f : \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ for Γ is periodic with period h (that is, invariant under T^h). Consequently, f admits a Fourier series expansion

$$f(z) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n z/h}.$$

We can then define

$$\widehat{f_\infty}(q_h) = \sum_{n=-\infty}^{\infty} a_n q_h^n$$

(where we have made the change $q_h = e^{2\pi i z/h}$) and we call it the q_h–expansion of f at infinity. This is because the map $z \mapsto q_h = e^{2\pi i z/h}$ induces an analytic isomorphism between $\langle T^h \rangle \backslash \mathbb{H}$ and the punctured disc of radius 1, and we can extend it to an isomorphism from $\langle T^h \rangle \backslash \mathbb{H}^*$ to the whole disc in such a way that $i\infty$ is mapped to 0. Now we say that f satisfies a certain property (is meromorphic or holomorphic, or vanishes) at ∞ if $\widehat{f_\infty}$ satisfies it at 0.

For a cusp $s \neq \infty$, we know that there exists $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ such that $s = \alpha(\infty)$. Let $g = f|_k^{[\alpha]}$, which is weakly modular for $\alpha^{-1}\Gamma\alpha$ (a congruence subgroup) because

$$g|_k^{[\alpha^{-1}\gamma\alpha]} = \left(f|_k^{[\alpha]}\right)\Big|_k^{[\alpha^{-1}\gamma\alpha]} = f|_k^{[\gamma\alpha]} = \left(f|_k^{[\gamma]}\right)\Big|_k^{[\alpha]} = f|_k^{[\alpha]} = g$$

for all $\alpha^{-1}\gamma\alpha \in \alpha^{-1}\Gamma\alpha$. We say that f satisfies a certain property (is meromorphic or holomorphic, or vanishes) at s if g satisfies the same property at ∞ (in the above sense). We define the width of s to be the minimum positive integer h such

that $\mathrm{T}^h \in \{\pm 1\} \cdot \alpha^{-1}\Gamma\alpha$. We call $\widehat{g_\infty}$ the q_h–expansion of f at s and write $\widehat{f_s} = \widehat{g_\infty}$.

Actually, we will only need to study the behaviour of functions in a set of representatives for the Γ–equivalence classes of cusps (and there are finitely many of them).

Proposition 1.13. *Let f be a weakly modular function for Γ. If $\alpha_1(\infty) = (\gamma\alpha_2)(\infty)$ for some $\alpha_1, \alpha_2 \in \mathrm{SL}_2(\mathbb{Z})$ and $\gamma \in \Gamma$, then the smallest power of q_h which occurs in the Fourier expansions of $f|_k^{[\alpha_1]}$ and $f|_k^{[\alpha_2]}$ is the same.*

Proof. Since $\alpha_1(\infty) = (\gamma\alpha_2)(\infty)$, necessarily $\alpha_1^{-1}\gamma\alpha_2 \in \mathrm{SL}_2(\mathbb{Z})_\infty = \langle \mathrm{T} \rangle$. That is, $\alpha_2 = \pm\gamma^{-1}\alpha_1\mathrm{T}^j$ for some integer j. Therefore,

$$f|_k^{[\alpha_2]} = (\pm 1)^k \left(f|_k^{[\alpha_1]} \right)\Big|_k^{[\mathrm{T}^j]} = (\pm 1)^k g|_k^{[\mathrm{T}^j]}$$

where $g = f|_k^{[\alpha_1]}$. If the q_h–expansion of g at ∞ is $g(z) = \sum_n a_n q_h^n$, the q_h–expansion

$$f|_k^{[\alpha_2]}(z) = (\pm 1)^k g(z + j) = (\pm 1)^k \sum_n a_n e^{2\pi i n j/\mathrm{N}} q_h^n$$

has the same non-zero coefficients (because the coefficients of the two series differ only by roots of unity). □

Definition 1.14. A *modular function* for Γ is a function $f : \mathbb{H}^* \to \mathbb{P}^1_{\mathbb{C}}$ satisfying the following conditions:

(i) f is invariant under the action of Γ on $\mathbb{H}^*$, i.e., $f \circ \gamma = f$ for all $\gamma \in \Gamma$;
(ii) f is meromorphic in $\mathbb{H}$;
(iii) f is meromorphic at the cusps.

Definition 1.15. A *modular form* for Γ of *weight k* is a function $f : \mathbb{H}^* \to \mathbb{P}^1_{\mathbb{C}}$ satisfying the following conditions:

(i) f is weakly modular for Γ of weight k;
(ii) f is holomorphic in $\mathbb{H}$;
(iii) f is holomorphic at the cusps.

If, in addition, f vanishes at all the cusps, we call it a *cusp form* for Γ of *weight k*. We refer to a function which satisfies (i) and also (ii) and (iii) with "holomorphic" replaced by "meromorphic" as a *meromorphic modular form* for Γ of *weight k*.

Despite the fact that we motivated the definition of modular forms as a means to obtain modular functions (through a relaxation of the required conditions), modular forms are interesting on their own right and have many interesting applications in number theory and several other areas of mathematics.

1.3 Eisenstein series, the modular discriminant and the modular invariant

In the remainder of this chapter, we focus on modular forms for the full modular group $\mathrm{SL}_2(\mathbb{Z})$.

The existence of (non-zero) modular forms is not obvious from definition 1.15. In this section we study some examples of modular forms of level 1. We observe that, since $-1 \in \mathrm{SL}_2(\mathbb{Z})$, there are no non-zero modular forms of odd weight (in that case, the condition of weak modularity for $\mathrm{SL}_2(\mathbb{Z})$ implies in particular that $f(z) = f(-1(z)) = -f(z)$). Therefore, we consider only modular forms of even weight $2k$ (for some integer k).

The examples exhibited in this section also play a fundamental role in the study of the spaces of modular forms for $\mathrm{SL}_2(\mathbb{Z})$.

Definition 1.16. Let $k > 1$. We define the *Eisenstein series* of index $2k$ as

$$\mathrm{G}_{2k}(z) = \sum_{m,n\in\mathbb{Z}}{}' \frac{1}{(mz+n)^{2k}}$$

for $z \in \mathbb{H}$. (The symbol $\sum'$ means that the summation runs over all values for which the corresponding addends "make sense"; in this case, over all pairs (m,n) distinct from $(0,0)$.)

Proposition 1.17. *If $k > 1$, the Eisenstein series G_{2k} converges to an holomorphic function on $\mathbb{H}$ which can be extended to a modular form for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$ with $\mathrm{G}_{2k}(\infty) = 2\zeta(2k)$, where ζ denotes the Riemann zeta function.*

Proof. Let $z \in \mathbb{H}$ and let $\mathrm{L}_z = \left\{ mz + n : (m,n) \in \mathbb{Z}^2 \setminus \{(0,0)\} \right\}$. For every $\mathrm{N} \in \mathbb{N}$, we consider the parallelogram P_N whose four vertices are the points $\pm\mathrm{N}z \pm \mathrm{N}$. If $r = \min\{|w| : w \in \mathrm{P}_1\}$, then $|w| \geq \mathrm{N}r$ for all $w \in \mathrm{P}_\mathrm{N}$. Since $\mathrm{P}_\mathrm{N} \cap \mathrm{L}_z$ contains exactly $8\mathrm{N}$ points for each N and $\mathrm{L}_z = \bigsqcup_{\mathrm{N}\in\mathbb{N}}(\mathrm{P}_\mathrm{N} \cap \mathrm{L}_z)$,

$$\sum_{m,n\in\mathbb{Z}}{}' |mz+n|^{-2k} = \sum_{\mathrm{N}>0} \sum_{w\in\mathrm{P}_\mathrm{N}\cap\mathrm{L}_z} |w|^{-2k} \leq \sum_{\mathrm{N}>0} 8\mathrm{N}(\mathrm{N}r)^{-2k} = 8r^{-2k} \sum_{\mathrm{N}>0} \mathrm{N}^{-2k+1}$$

and this last series converges because $-2k+1 < -1$. This proves that G_{2k} converges absolutely.

Now we check that G_{2k} defines a weakly modular function of weight $2k$.

Indeed, we recall that $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ generate $SL_2(\mathbb{Z})$ and compute

$$G_{2k}(z+1) = \sum_{m,n\in\mathbb{Z}}{}' (mz+m+n)^{-2k} = \sum_{m,n\in\mathbb{Z}}{}' (mz+n)^{-2k} = G_{2k}(z),$$
$$G_{2k}\left(\frac{-1}{z}\right) = \sum_{m,n\in\mathbb{Z}}{}' \left(\frac{m}{z}+n\right)^{-2k} = z^{2k} \sum_{m,n\in\mathbb{Z}}{}' (m+nz)^{-2k} = z^{2k} G_{2k}(z).$$

Let F be the fundamental domain for $SL_2(\mathbb{Z})$ defined in theorem 1.9. If $z \in F$, we have that

$$|mz+n|^2 = m^2|z|^2 + 2mn\mathfrak{Re}(z) + n^2 \geq m^2 - mn + n^2 = |m\rho - n|^2$$

where $\rho = e^{2\pi i/3}$. But we have already proved that $G_{2k}(\rho) = \sum'(m\rho - n)^{-2k}$ converges absolutely. This means that $G_{2k}(z)$ converges uniformly in F, and thus also in each of the transforms γF (applying the result to $G_{2k}(\gamma^{-1}(z))$). Since these cover $\mathbb{H}$, we conclude that $G_{2k}(z)$ converges uniformly absolutely on compact subsets of $\mathbb{H}$. In particular, G_{2k} defines a weakly modular function which is holomorphic in $\mathbb{H}$.

It remains to see that G_{2k} is holomorphic at infinity. To this aim, we need to prove that G_{2k} has a limit for $\mathfrak{Im}(z) \to \infty$. But one may suppose that z remains in F. Since G_{2k} converges uniformly in F, we can make the passage to the limit term by term:

$$\lim_{z\to i\infty} G_{2k}(z) = \sum_{m,n\in\mathbb{Z}}{}' \lim_{z\to i\infty} \frac{1}{(mz+n)^{2k}} = \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n^{2k}} = 2\sum_{n=1}^{\infty} \frac{1}{n^{2k}} = 2\zeta(2k)$$

(here, we used that the terms with $m \neq 0$ tend to 0). □

Proposition 1.18. *For every integer $k > 1$, the q–expansion of the Eisenstein series G_{2k} is*

$$G_{2k}(z) = 2\zeta(2k) + 2\frac{(2\pi i)^{2k}}{(2k-1)!} \sum_{n=1}^{\infty} \sigma_{2k-1}(n) q^n$$

where

$$\sigma_j(n) = \sum_{d|n} d^j$$

is the sum of j–th powers of positive divisors of n and $q = q(z) = e^{2\pi i z}$.

Proof. We start with the well-known formula

$$\frac{1}{z}+\sum_{n=1}^{\infty}\left(\frac{1}{z+n}+\frac{1}{z-n}\right)=\pi\cot(\pi z)=\pi\frac{\cos(\pi z)}{\sin(\pi z)}=\pi i-\frac{2\pi i}{1-q}=\pi i-2\pi i\sum_{r=0}^{\infty}q^r,$$

which can be obtained taking the logarithmic derivative of the expression of $\sin(\pi z)$ as an infinite product. By successive differentiations, we obtain the formula

$$\sum_{n\in\mathbb{Z}}\frac{1}{(z+n)^j}=\frac{1}{(j-1)!}(-2\pi i)^j\sum_{r=1}^{\infty}r^{j-1}q^r,$$

for $j\geq 2$. After replacing z with mz, this becomes

$$\sum_{n\in\mathbb{Z}}\frac{1}{(mz+n)^j}=\frac{1}{(j-1)!}(-2\pi i)^j\sum_{r=1}^{\infty}r^{j-1}q^{mr}.$$

Finally, we use this to expand

$$\begin{aligned}\mathrm{G}_{2k}(z)&=\sum_{m,n\in\mathbb{Z}}{}'\frac{1}{(mz+n)^{2k}}=2\zeta(2k)+2\sum_{m=1}^{\infty}\sum_{n\in\mathbb{Z}}\frac{1}{(mz+n)^{2k}}\\&=2\zeta(2k)+\frac{2(-2\pi i)^{2k}}{(2k-1)!}\sum_{m=1}^{\infty}\sum_{r=1}^{\infty}r^{2k-1}q^{mr}=2\zeta(2k)+\frac{2(2\pi i)^{2k}}{(2k-1)!}\sum_{n=1}^{\infty}\sigma_{2k-1}(n)\cdot q^n\end{aligned}$$

as required. □

These explicit q–expansions can be used to derive identities relating the arithmetic functions $\sigma_j(n)$.

We have seen that there exists a family of non-zero modular forms, and we have even computed their q–expansions. We next use Eisenstein series to define a cusp form and a modular function.

Definition 1.19. We define the functions $g_4(z)=60\mathrm{G}_4(z)$ and $g_6(z)=140\mathrm{G}_6(z)$ (it is convenient to choose these multiples of the corresponding Eisensten series because of their relation to the theory of elliptic curves).

Definition 1.20. The *modular discriminant* is the cusp form for $\mathrm{SL}_2(\mathbb{Z})$ of weight 12 defined by

$$\Delta(z)=g_4(z)^3-27g_6(z)^2.$$

One checks that the constants in this definition are chosen so that $\Delta(\infty)=0$. In addition, $\Delta(z)$ is a modular form of weight 12 because the product of modular

forms is a modular form (and its weight is the sum of weights). In conclusion, $\Delta(z)$ is a cusp form of weight 12.

The modular discriminant has also been studied for its arithmetic properties (the coefficients of its q–expansion define the Ramanujan τ–function up to a constant factor). We shall see that it is in fact the (non-zero) cusp form of least possible weight. Its q–expansion can be computed using the following result, which we state without proof (see theorem 6 of chapter VII of Serre's book [11] for an elementary proof).

Theorem 1.21 (Jacobi). *The modular discriminant's can be expressed as*

$$\Delta(z) = (2\pi)^{12} q \prod_{n=1}^{\infty} (1-q^n)^{24}, \quad q = q(z) = e^{2\pi i z}.$$

Corollary 1.22. $\Delta(z)$ *does not vanish in* $\mathbb{H}$ *and has a simple zero at* $i\infty$.

We shall give a different proof of this corollary later.

As commented before, modular functions can be obtained as quotients of modular forms.

Definition 1.23. The *modular invariant* is the modular function for $\mathrm{SL}_2(\mathbb{Z})$

$$j(z) = 1728 \frac{g_4(z)^3}{\Delta(z)} = 1728 \frac{g_4(z)^3}{g_4(z)^3 - 27 g_6(z)^2}.$$

It is clear from the definition that $j(z)$ is a modular function for $\mathrm{SL}_2(\mathbb{Z})$. In particular, it is holomorphic in $\mathbb{H}$ and has a simple pole at $i\infty$. The coefficient $1728 = 2^6 3^3$ has been introduced in order that $j(z)$ has a residue equal to 1 at infinity. Furthermore, an interesting property of j is that it defines an analytic isomorphism between $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}^*$ and the Riemann sphere $\mathbb{C} \cup \{\infty\}$.

1.4 Structure theorem

After seeing the examples of the previous section, we are in a position to describe explicitly the structure of the space of modular forms for $\mathrm{SL}_2(\mathbb{Z})$.

Lemma 1.24. *Let* Γ *be a congruence subgroup of* $\mathrm{SL}_2(\mathbb{Z})$ *and let* k *be an integer. The set* $\mathrm{M}_k(\Gamma)$ *of modular forms for* Γ *of weight* k *is a vector space over* $\mathbb{C}$. *Furthermore, the set* $\mathrm{S}_k(\Gamma)$ *of cusp forms for* Γ *of weight* k *is a subspace of finite codimension in* $\mathrm{M}_k(\Gamma)$.

Proof. On the one hand, sums and scalar multiples of holomorphic functions are again holomorphic. Moreover, the right action of $\mathrm{GL}_2(\mathbb{R})$ on functions defined on $\mathbb{H}^*$ commutes with linear combinations and, consequently, the condition of being weakly modular of weight k is also satisfied by sums and scalar multiples of weakly modular functions of weight k. The zero function is a modular form of weight k too. In conclusion, $\mathrm{M}_k(\Gamma)$ is a vector space over $\mathbb{C}$.

On the other hand, consider a set C of representatives of the Γ–equivalence classes of cusps. The kernel of the linear map $f \mapsto (f(s))_{s\in \mathrm{C}} : \mathrm{M}_k(\Gamma) \to \mathbb{C}^{|\mathrm{C}|}$ is precisely $\mathrm{S}_k(\Gamma)$. Therefore, $\mathrm{S}_k(\Gamma)$ is a subspace of $\mathrm{M}_k(\Gamma)$ and, by the first isomorphism theorem, $\operatorname{codim}(\mathrm{S}_k(\Gamma)) \leq \dim(\mathbb{C}^{|\mathrm{C}|}) = |\mathrm{C}| < \infty$. □

Proposition 1.25. *Let $k > 1$. The subspace $\mathrm{S}_{2k}(\mathrm{SL}_2(\mathbb{Z}))$ has codimension 1 in $\mathrm{M}_{2k}(\mathrm{SL}_2(\mathbb{Z}))$ and $\mathrm{M}_{2k}(\mathrm{SL}_2(\mathbb{Z})) = \mathrm{S}_{2k}(\mathrm{SL}_2(\mathbb{Z})) \oplus \mathbb{C}\mathrm{G}_{2k}$.*

Proof. This is a consequence of lemma 1.24. Indeed, all the cusps are equivalent under $\mathrm{SL}_2(\mathbb{Z})$ and the Eisenstein series G_{2k} is an element of M_{2k} such that $\mathrm{G}_{2k}(i\infty) \neq 0$. Therefore, $\mathrm{M}_{2k}(\mathrm{SL}_2(\mathbb{Z})) = \mathrm{S}_{2k}(\mathrm{SL}_2(\mathbb{Z})) \oplus \mathbb{C}\mathrm{G}_{2k}$. □

The last result gives us the structure of the space of modular forms for $\mathrm{SL}_2(\mathbb{Z})$ of a given weight. We shall find explicit bases of these vector spaces, but first we study how they relate to each other.

Proposition 1.26. *Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. The set of modular forms for Γ is an associative, commutative and unital graded algebra over $\mathbb{C}$:*

$$\mathrm{M}(\Gamma) = \bigoplus_{k\in\mathbb{Z}} \mathrm{M}_k(\Gamma).$$

Proof. This result follows from lemma 1.24 and the fact that, if f and g are modular forms for Γ of weights k and l, respectively, the product fg is a modular form for Γ of weight $k + l$. Indeed, the product of holomorphic functions is holomorphic and

$$(fg)(\gamma(z)) = f(\gamma(z))g(\gamma(z)) = (cz+d)^k f(z)(cz+d)^l g(z) = (cz+d)^{k+l}(fg)(z)$$

for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ and all $z \in \mathbb{H}^*$. □

We can now state and prove the main results, which allow us to make explicit computations with modular forms for $\mathrm{SL}_2(\mathbb{Z})$. That is to say, our aim is to find a basis of these vector spaces.

Definition 1.27. Let f be a meromorphic modular form for a congruence subgroup Γ, not identically zero. The *order* of f at a point $p \in \mathbb{H}$ is the integer $\operatorname{ord}_p(f)$ such that $f(z)(z-p)^{-\operatorname{ord}_p(f)}$ is holomorphic and non-zero at p. The *order* of f at a cusp s is the integer $\operatorname{ord}_s(f)$ such that $\widehat{f_s}(q_h)q_h^{-\operatorname{ord}_s(f)}$ is holomorphic and non-zero at 0, where $\widehat{f_s}$ is the q_h–expansion of f at s (and h is the width of s).

Theorem 1.28 (valence formula). *Let k be an integer and let f be a meromorphic modular form for $\mathrm{SL}_2(\mathbb{Z})$ of weight k, not identically zero. Then,*

$$\operatorname{ord}_\infty(f) + \frac{1}{2}\operatorname{ord}_i(f) + \frac{1}{3}\operatorname{ord}_\rho(f) + \sum_p{}^{*} \operatorname{ord}_p(f) = \frac{k}{12},$$

where $\sum_p^*$ means a summation over the points of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ distinct from the classes of i and ρ.

Proof. Let F be the fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$ described in theorem 1.9.

Observe that the sum we want to compute "makes sense" (i.e., that f has a finite number of zeros and poles modulo $\mathrm{SL}_2(\mathbb{Z})$). Indeed, since f is meromorphic at infinity, there exists some $R > 0$ such that f has neither zeros nor poles in $\{z \in \mathbb{H} : \operatorname{Im}(z) \geq R\}$. That is, the q–expansion $\widehat{f_\infty}$ of f at infinity is meromorphic at 0 and so there exists some $r > 0$ such that $\widehat{f_\infty}$ has neither zeros nor poles in the punctured disc $\{q : 0 < |q| \leq r\}$. Here, $r = e^{-2\pi R}$. And $\mathrm{F}_R = \{z \in \mathrm{F} : \operatorname{Im}(z) \leq R\}$ contains only a finite number of zeros and poles because it is compact and f is meromorphic in $\mathbb{H}$.

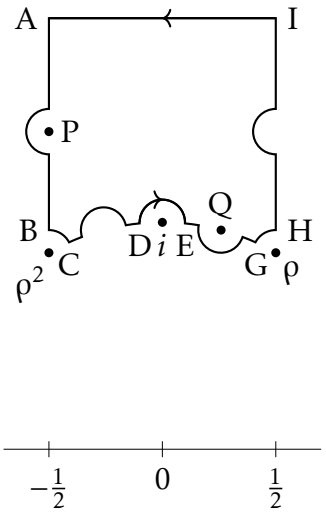


Figure 1.2: Contour for the proof of theorem 1.28.

The zeros and poles of f are all equivalent to either some point in F or to ∞. The main idea of the proof is to integrate the logarithmic derivative of f round a contour (which is similar to the boundary of F). More precisely, let $\mathcal{C}$ be the contour depicted in figure 1.2: we are going to integrate the meromorphic function $\frac{f'}{f}$ round $\mathcal{C}$. The top of $\mathcal{C}$ is a horizontal line from $\mathrm{I} = \frac{1}{2}+i\mathrm{R}$ to $\mathrm{A} = -\frac{1}{2}+i\mathrm{R}$. The rest of the contour follows round the boundary of the fundamental domain F, except that it detours round any zero or pole on the boundary along circular arcs of small radius ε (we are going to take the limit as ε approaches 0). This is done in such a way as to include every equivalence class of zero or pole exactly once inside $\mathcal{C}$ other than i and ρ (and so $\rho^2 = \mathrm{S}(\rho)$), which are left outside of $\mathcal{C}$ if they are zeros or poles.

Figure 1.2 illustrates the case in which the zeros and poles on the boundary of F (i.e., the points which have to be avoided when defining the contour $\mathcal{C}$) are precisely the points i, ρ and ρ^2, one point P on the vertical boundary (and its equivalent point on the opposite line) and one point Q on the unit circle (and its equivalent point also on the unit circle).

By the residue theorem,

$$\frac{1}{2\pi i}\int_{\mathcal{C}} \frac{f'(z)}{f(z)}\,dz = \sum_p{}^{*} \operatorname{ord}_p(f).$$

We evaluate this integral dividing the contour $\mathcal{C}$ in parts. (See figure 1.2 for the points used in the division of the contour.)

First of all, the integral from A to B cancels the integral from H to I because $f(z+1) = f(z)$ (and the lines go in opposite directions).

To evaluate the integral from I to A, we consider the change of variables $q = e^{2\pi i z}$. This section of the integral is thus equal to the integral

$$\frac{1}{2\pi i}\int \frac{\widehat{f}'_\infty(q)}{\widehat{f}_\infty(q)}\,dq$$

along the circle of radius $r = e^{-2\pi \mathrm{R}}$ centred at the origin with negative orientation. Therefore, the value of this integral is $-\operatorname{ord}_\infty(f)$ (i.e., $-\operatorname{ord}_0(\widehat{f}_\infty)$).

The integral along the circle which contains DE, oriented negatively, has the value $-\operatorname{ord}_i(f)$. And, as ε tends to 0, the angle $\angle \mathrm{D}i\mathrm{E}$ tends to π. Hence,

$$\frac{1}{2\pi i}\int_{\mathrm{D}}^{\mathrm{E}} \frac{f'(z)}{f(z)}\,dz \xrightarrow[\varepsilon\to 0^+]{} -\frac{\operatorname{ord}_i(f)}{2}.$$

Similarly,

$$\frac{1}{2\pi i}\int_B^C \frac{f'(z)}{f(z)}\,dz \xrightarrow[\varepsilon\to 0^+]{} -\frac{\mathrm{ord}_\rho(f)}{6} \quad \text{and} \quad \frac{1}{2\pi i}\int_G^H \frac{f'(z)}{f(z)}\,dz \xrightarrow[\varepsilon\to 0^+]{} -\frac{\mathrm{ord}_\rho(f)}{6}.$$

Finally, we observe that $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ transforms the path from C to D into the path from G to E. Since $f(S(z)) = z^k f(z)$,

$$\frac{1}{2\pi i}\left(\int_C^D \frac{f'(z)}{f(z)}\,dz + \int_E^G \frac{f'(z)}{f(z)}\,dz\right) = \frac{1}{2\pi i}\int_C^D \left[\frac{f'(z)}{f(z)} - \frac{f'(S(z))}{f(S(z))}\right]dz = \frac{1}{2\pi i}\int_C^D -k\frac{dz}{z}$$

and this last integral tends to $\frac{k}{12}$ as ε tends to 0 because $\angle C0D$ tends to $\frac{\pi}{6}$. □

We can use this theorem to prove a result which was stated before and which we are going to need.

Corollary 1.22. *$\Delta(z)$ does not vanish in $\mathbb{H}$ and has a simple zero at $i\infty$.*

Proof. Since Δ is a cusp form of weight 12, $\mathrm{ord}_p(\Delta) \geq 0$ for all $p \in \mathbb{H} \setminus \mathrm{SL}_2(\mathbb{Z})$ and $\mathrm{ord}_\infty(\Delta) \geq 1$. By theorem 1.28, these numbers add up to 1: this is only possible if all the inequalities are equalities. □

Proposition 1.29. *Let k be an integer. Multiplication by Δ defines an isomorphism of $M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$ onto $S_k(\mathrm{SL}_2(\mathbb{Z}))$.*

Proof. Clearly, if $f \in M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$, then $f\Delta \in S_k(\mathrm{SL}_2(\mathbb{Z}))$. For the converse, let $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$. We set $g = \frac{f}{\Delta}$, which is a meromorphic modular form of weight $k-12$. Using the previous result, we obtain that $\mathrm{ord}_p(g) = \mathrm{ord}_p(f) \geq 0$ for every $p \in \mathbb{H}$ and $\mathrm{ord}_\infty(g) = \mathrm{ord}_\infty(f) - 1 \geq 0$. In conclusion, $g \in M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$.

This proves that multiplication by Δ gives a bijection between $M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$ and $S_k(\mathrm{SL}_2(\mathbb{Z}))$, and it is obviously a linear transformation. □

Proposition 1.30. *Let k be an integer.*

(1) $M_k(\mathrm{SL}_2(\mathbb{Z})) = 0$ *if $k < 0$, k is odd or $k = 2$.*
(2) $M_0(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}$ *(that is, the only modular forms for $\mathrm{SL}_2(\mathbb{Z})$ of weight 0 are the constants).*
(3) $M_k(\mathrm{SL}_2(\mathbb{Z}))$ *has dimension 1 and G_k is a basis if $k = 4, 6, 8, 10$ or 14.*

Proof. Recall that, if f is a non-zero modular form of weight k, then

$$\operatorname{ord}_\infty(f) + \frac{1}{2}\operatorname{ord}_i(f) + \frac{1}{3}\operatorname{ord}_\rho(f) + \sum_p{}^{*} \operatorname{ord}_p(f) = \frac{k}{12}.$$

Since f is holomorphic, this means that $\frac{k}{12} \geq 0$. Moreover, k must be even because the least common denominator of the left-hand side is 6 and $k \neq 2$ because $\frac{1}{6}$ cannot be written in the form $a + \frac{b}{2} + \frac{c}{3}$ with $a, b, c \geq 0$.

If $k \leq 10$, we have that $k - 12 < 0$ and $S_k(\mathrm{SL}_2(\mathbb{Z})) = \{0\}$ by proposition 1.29. Therefore, $\dim(M_k(\mathrm{SL}_2(\mathbb{Z}))) \leq 1$. But we already know that $1, G_4, G_6, G_8, G_{10}$ are non-zero modular forms for $\mathrm{SL}_2(\mathbb{Z})$ of weights $0, 4, 6, 8, 10$, respectively; this concludes the proof. □

Corollary 1.31. *For any integer k,*

$$\dim(M_k(\mathrm{SL}_2(\mathbb{Z}))) = \begin{cases} 0 & \text{if } k < 0,\ k \text{ is odd or } k = 2, \\ \left\lfloor \frac{k}{12} \right\rfloor & \text{if } k \geq 0,\ k \text{ is even and } k \equiv 1 \pmod 6, \\ \left\lfloor \frac{k}{12} \right\rfloor + 1 & \text{if } k \geq 0,\ k \text{ is even and } k \not\equiv 1 \pmod 6. \end{cases}$$

Proof. The result follows by induction on k (the inductive step is performed by increasing k to $k + 12$ using proposition 1.29). □

Theorem 1.32. *Let $k \geq 0$. The vector space $M_{2k}(\mathrm{SL}_2(\mathbb{Z}))$ admits as a basis the family of monomials $G_4^\alpha G_6^\beta$ with α and β non-negative integers such that $2\alpha + 3\beta = k$. As a consequence, $M(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}[G_4, G_6]$.*

Proof. First, we show that these monomials generate $M_{2k}(\mathrm{SL}_2(\mathbb{Z}))$ by induction on k. This is clear for $k \leq 3$, so suppose that $k \geq 4$. Choose a pair (α_0, β_0) of non-negative integers such that $2\alpha_0 + 3\beta_0 = k$ (this is possible for $k \geq 2$). The modular form $g = G_4^{\alpha_0} G_6^{\beta_0}$, of weight $2k$, is not a cusp form. Let $f \in M_{2k}(\mathrm{SL}_2(\mathbb{Z}))$. Now $f - \frac{f(\infty)}{g(\infty)} g$ is a cusp form and, in particular, is of the form Δh for some $h \in M_{2k-12}(\mathrm{SL}_2(\mathbb{Z}))$. We can apply the induction hypothesis to h and obtain thus f as a linear combination of the desired monomials.

Now we see that these monomials are linearly independent. Suppose, for the sake of contradiction, that there exists a non-trivial linear combination

$$\sum_{2\alpha+3\beta=k} \lambda_{\alpha,\beta} G_4^\alpha G_6^\beta = 0.$$

Up to multiplying this linear relation by suitable powers of G_4 and of G_6, we may assume that k is a multiple of 12. Dividing by $G_6^{k/3}$, we obtain that

$$\sum_{2\alpha+3\beta=k} \lambda_{\alpha,\beta}\left(\frac{G_4^3}{G_6^2}\right)^{\frac{\alpha}{3}} = 0$$

(where $\frac{\alpha}{3}$ is an integer because $2\alpha = k - 3\beta$ is a multiple of 3). That is, the meromorphic function G_4^3/G_6^2 satisfies a non-trivial algebraic equation over $\mathbb{C}$ and, therefore, is constant. But this is not the case: $G_4(\rho) = 0 \neq G_6(\rho)$ whereas $G_4(i) \neq 0 = G_6(i)$. □

Hence, we know a basis of each space of modular forms for $\mathrm{SL}_2(\mathbb{Z})$ and we can compute the q–expansions of the elements of this basis explicitly. There is a similar result for modular functions for $\mathrm{SL}_2(\mathbb{Z})$.

Proposition 1.33. *The modular invariant j defines by passage to quotient a bijection of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}^*$ onto $\mathbb{P}^1_{\mathbb{C}}$.*

Proof. We already know that j has a simple pole at infinity and is holomorphic in $\mathbb{H}$. Thus, we have to prove that the modular form $f_\lambda = 1728g_2^3 - \lambda\Delta$ has a unique zero modulo $\mathrm{SL}_2(\mathbb{Z})$ for all $\lambda \in \mathbb{C}$ (recall that $j = 1728g_2^3/\Delta$). By theorem 1.28,

$$\mathrm{ord}_\infty(f_\lambda) + \frac{1}{2}\mathrm{ord}_i(f_\lambda) + \frac{1}{3}\mathrm{ord}_\rho(f_\lambda) + \sum_p{}^* \mathrm{ord}_p(f_\lambda) = 1$$

and the only decompositions of 1 in the form $a + \frac{b}{2} + \frac{c}{3}$ with $a, b, c \geq 0$ correspond to $(a, b, c) = (1, 0, 0)$, $(0, 2, 0)$ or $(0, 0, 3)$. In all three cases f_λ vanishes at exactly one point of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. □

Actually, this bijection is an isomorphism of Riemann surfaces (with the structure of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}^*$ defined in the next chapter).

Theorem 1.34. *Let f be a meromorphic function on $\mathbb{H}^*$. The following properties are equivalent:*

(a) *f is a modular function for $\mathrm{SL}_2(\mathbb{Z})$;*
(b) *f is a quotient of two modular forms for $\mathrm{SL}_2(\mathbb{Z})$ of the same weight;*
(c) *f is a rational function of j.*

Proof. The implications (c) $\Longrightarrow$ (b) $\Longrightarrow$ (a) are immediate.

Now we prove that (a) $\Longrightarrow$ (c). After multiplying f by a suitable polynomial in j, we may assume that f is holomorphic in $\mathbb{H}$. Since Δ is a cusp form, there exists an integer $n \geq 0$ such that $g = \Delta^n f$ is holomorphic at infinity as well. Thus g is a modular form of weight $12n$ and can be expressed as a linear combination of monomials $G_4^\alpha G_6^\beta$ with $2\alpha + 3\beta = 6n$. By linearity, we are reduced to the case $g = G_4^\alpha G_6^\beta$. That is,

$$f = \frac{G_4^\alpha G_6^\beta}{\Delta^n} = \left(\frac{G_4^3}{\Delta}\right)^{\frac{\alpha}{3}} \left(\frac{G_6^2}{\Delta}\right)^{\frac{\beta}{2}} = \left(\frac{j}{1728 \cdot 60^3}\right)^{\frac{\alpha}{3}} \left(\frac{j}{1728 \cdot 27 \cdot 140^2} - \frac{1}{27 \cdot 140^2}\right)^{\frac{\beta}{2}}$$

and we observe that both $\frac{\alpha}{3}$ and $\frac{\beta}{2}$ are integers. $\square$

Chapter 2

Modular curves as Riemann surfaces

The set of orbits of $\mathbb{H}$ under the action of a congruence subgroup can be endowed with the structure of a Riemann surface. This chapter describes the construction of this Riemann surface, which is called modular curve, and explains how it can be compactified by adding a finite number of points. In some sense, some of the facts presented in chapter 1 in relation to the fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$ are generalised here using results from the theory of Riemann surfaces.

The exposition of this chapter is based principally on some sections of the first half of Milne's notes [8] as well as on the relevant chapters of the books [12] by Shimura, [4] by Lang and [2] by Diamond and Shurman. (The first chapter of Shimura's book [12] gives a complete description of modular curves for more general types of groups.) Also, Reyssat's book [10] develops the theory of Riemann surfaces and devotes a whole chapter to examples related to modular curves.

2.1 Classification of Möbius transformations

Consider the action of $\mathrm{GL}_2(\mathbb{C})$ on $\mathbb{P}^1_{\mathbb{C}}$ given by linear fractional transformations. The scalar matrices $\left(\begin{smallmatrix}\lambda & 0\\ 0 & \lambda\end{smallmatrix}\right)$ act as the identity transformation. Let $\alpha \in \mathrm{GL}_2(\mathbb{C})$ and suppose that it is not a scalar matrix. By the theory of Jordan canonical forms, α is conjugate to a matrix of one of the following two forms:

$$\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix}\lambda & 0\\ 0 & \mu\end{pmatrix},\ \lambda \neq \mu.$$

In the first case, α is conjugate to a transformation $z \mapsto z + \lambda^{-1}$; in the second case, to a transformation $z \mapsto cz$, $c \neq 1$.

Definition 2.1. A non-scalar matrix $\alpha \in \mathrm{GL}_2(\mathbb{C})$ (or the corresponding non-trivial linear fractional transformation) is called

(1) *parabolic* if it is conjugate to a transformation of the form $z \mapsto z + \lambda^{-1}$,

(2) *elliptic* if it is conjugate to a transformation of the form $z \mapsto cz$ with $|c| = 1$,

(3) *hyperbolic* if it is conjugate to a transformation of the form $z \mapsto cz$ with $c \in \mathbb{R}^+$, and
(4) *loxodromic* otherwise.

A non-scalar $\alpha \in \mathrm{GL}_2(\mathbb{C})$ has one or two fixed points in $\mathbb{P}^1_{\mathbb{C}}$ depending on whether it is parabolic or not. We will be interested in the fixed points of the elements of $\mathrm{SL}_2(\mathbb{R})$. In this case, the classification of linear fractional transformations becomes simpler.

Proposition 2.2. *Let* $\alpha \in \mathrm{SL}_2(\mathbb{C}) \setminus \{\pm 1\}$.

(1) α *is parabolic if and only if* $\mathrm{tr}(\alpha) = \pm 2$.
(2) α *is elliptic if and only if* $\mathrm{tr}(\alpha)$ *is real and* $|\mathrm{tr}(\alpha)| < 2$.
(3) α *is hyperbolic if and only if* $\mathrm{tr}(\alpha)$ *is real and* $|\mathrm{tr}(\alpha)| > 2$.
(4) α *is loxodromic if and only if* $\mathrm{tr}(\alpha)$ *is not real.*

Proof. Since $\det(\alpha) = 1$, the Jordan canonical form for α is either $\left(\begin{smallmatrix} \pm 1 & 1 \\ 0 & \pm 1 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{smallmatrix}\right)$, $\lambda \neq \pm 1$. (1) follows immediately from this.

If α is elliptic, then $c = \lambda^2$ and $|c| = 1$, so $\mathrm{tr}(\alpha) = \lambda + \lambda^{-1} = 2\Re\mathfrak{e}(\lambda)$ is real and its absolute value is < 2 (because $\lambda \neq \pm 1$). Similarly, if α is hyperbolic, then $c = \lambda^2 \in \mathbb{R}^+$, so $\lambda \in \mathbb{R}^+ \setminus \{0, \pm 1\}$; consequently, $\mathrm{tr}(\alpha) = \lambda + \lambda^{-1}$ is real and its absolute value is > 2.

Conversely, suppose that α is conjugate to $\left(\begin{smallmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{smallmatrix}\right)$ and that $\mathrm{tr}(\alpha) = \lambda + \lambda^{-1}$ is real. If λ is real, since $\lambda \neq \pm 1$, we obtain that $|\mathrm{tr}(\alpha)| > 2$ and α must be hyperbolic because $c = \lambda^2 \in \mathbb{R}^+$. Otherwise, λ and $\overline{\lambda}$ are the roots of $X^2 - \mathrm{tr}(\alpha)X + 1 = 0$; therefore, $\lambda\overline{\lambda} = 1$ and α must be elliptic. □

It is clear that there are no loxodromic elements in $\mathrm{SL}_2(\mathbb{R})$. Let us study the behaviour of the other types of transformations.

If $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R}) \setminus \{\pm 1\}$ is parabolic, it has exactly one eigenvector (up to scalar multiplication), which is real. Let $\left(\begin{smallmatrix} e \\ f \end{smallmatrix}\right)$ be this eigenvector. If $f \neq 0$, α has a fixed point $z = \frac{e}{f}$ in $\mathbb{R}$ which is actually the double root of $\frac{az+b}{cz+d} = z$. If, contrariwise, $f = 0$, then $c = 0$ and $\alpha = \pm\left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right)$; thus, its only fixed point is ∞.

If $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R}) \setminus \{\pm 1\}$ is elliptic, its characteristic polynomial is $X^2 - (a+d)X + 1$ with $|a+d| < 2$. Therefore, α has two complex conjugate eigenvectors $\left(\begin{smallmatrix} e \\ f \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} \overline{e} \\ \overline{f} \end{smallmatrix}\right)$ which correspond to two complex conjugate fixed points $z = \frac{e}{f}$ and $\overline{z} = \frac{\overline{e}}{\overline{f}}$ (one of them in $\mathbb{H}$).

If $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R}) \setminus \{\pm 1\}$ is hyperbolic, its characteristic polynomial is $X^2 - (a+d)X + 1$ with $|a+d| > 2$. Therefore, α has two linearly independent real eigenvectors which correspond to two distinct fixed points in $\mathbb{R} \cup \{\infty\}$.

Proposition 2.3. *Let $\alpha \in \mathrm{SL}_2(\mathbb{R}) \setminus \{\pm 1\}$.*

(1) *α is parabolic if and only if it has exactly one fixed point in $\mathbb{R} \cup \{\infty\}$.*
(2) *α is elliptic if and only if it has exactly one fixed point z in $\mathbb{H}$ (and the other fixed point is $\overline{z}$).*
(3) *α is hyperbolic if and only if it has two distinct fixed points in $\mathbb{R} \cup \{\infty\}$.*

Corollary 2.4. *Let $\alpha \in \mathrm{SL}_2(\mathbb{R}) \setminus \{\pm 1\}$ and let $m \in \mathbb{Z}$ such that $\alpha^m \neq \pm 1$. α is parabolic (resp. elliptic or hyperbolic) if and only if α^m is parabolic (resp. elliptic or hyperbolic).*

We now classify the points of $\mathbb{H}$ and $\mathbb{P}^1_{\mathbb{R}}$ with respect to the action of a fixed discrete subgroup Γ of $\mathrm{SL}_2(\mathbb{R})$.

Definition 2.5. Consider the action of Γ given by Möbius transformations.

(1) A point $z \in \mathbb{H}$ is called an *elliptic point* for Γ if it is the fixed point of an elliptic element γ of Γ.
(2) A point $s \in \mathbb{P}^1_{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ is called a *cusp* for Γ if it is the fixed point of a parabolic element γ of Γ.

Proposition 2.6. *If z is an elliptic point for Γ, then Γ_z is a cyclic group.*

Proof. If $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$, $\alpha(i) = i$ if and only if $ai + b = di - c$. Therefore,

$$\mathrm{SL}_2(\mathbb{R})_i = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) : a = d,\, b = -c \text{ and } a^2 + b^2 = 1 \right\} = \mathrm{SO}_2(\mathbb{R}).$$

Since $\mathrm{SL}_2(\mathbb{R})$ acts transitively on $\mathbb{H}$, there exists some $\sigma \in \mathrm{SL}_2(\mathbb{R})$ such that $\sigma(i) = z$. In this case, $\mathrm{SL}_2(\mathbb{R})_z = \sigma\,\mathrm{SO}_2(\mathbb{R})\sigma^{-1}$. Hence, $\Gamma_z = \sigma\,\mathrm{SO}_2(\mathbb{R})\sigma^{-1} \cap \Gamma$ and this group is finite because Γ is discrete and $\mathrm{SO}_2(\mathbb{R})$ is compact. Finally, we observe that $\mathrm{SO}_2(\mathbb{R})$ is isomorphic to $\mathbb{R}/\mathbb{Z}$ and every finite subgroup of $\mathbb{R}/\mathbb{Z}$ is cyclic (in fact, every finite subgroup of $\mathbb{R}/\mathbb{Z}$ is of the form $n^{-1}\mathbb{Z}/\mathbb{Z}$ where n is the least common denominator of the elements of the subgroup). □

Proposition 2.7. *The elements of Γ of finite order are precisely the elliptic elements of Γ and ± 1.*

Proof. Let $\alpha \in \mathrm{SL}_2(\mathbb{R})$ and suppose that it has finite order. By the theory of Jordan canonical forms, α is conjugate in $\mathrm{SL}_2(\mathbb{C})$ to a matrix of the form $\left(\begin{smallmatrix} \zeta & 0 \\ 0 & \overline{\zeta} \end{smallmatrix}\right)$, where ζ is a root of unity. In this case, $\mathrm{tr}(\alpha) = \zeta + \overline{\zeta} = 2\mathfrak{Re}(\zeta)$. Therefore, $|\mathrm{tr}(\alpha)| < 2$ and α is elliptic unless $\zeta = \pm 1$. The converse is a consequence of proposition 2.6. □

Proposition 2.8. *Let Γ_1 and Γ_2 be two discrete subgroups of $\mathrm{SL}_2(\mathbb{R})$. If $\Gamma_1 \cap \Gamma_2$ is a subgroup of finite index in both Γ_1 and Γ_2, then Γ_1 and Γ_2 have the same set of cusps.*

Proof. Up to replacing Γ_2 with $\Gamma_1 \cap \Gamma_2$, we may assume that $\Gamma_2 \subseteq \Gamma_1$. It is clear that a cusp for Γ_2 is also a cusp for Γ_1 because the parabolic elements of Γ_2 are also parabolic elements of Γ_1.

If s is a cusp for Γ_1, then $\gamma(s) = s$ for some parabolic element of Γ_1. Since $[\Gamma_1 : \Gamma_2]$ is finite, $\gamma\Gamma_2$ has finite order in Γ_1/Γ_2. That is to say, there exists a positive integer n such that $\gamma^n \in \Gamma_2$. But $\gamma^n(s) = s$ and γ^n is also parabolic by corollary 2.4. (Note that proposition 2.7 ensures that $\gamma^n \neq \pm 1$.) □

Example 2.9. The cusps for the full modular group $\mathrm{SL}_2(\mathbb{Z})$ are exactly the elements of $\mathbb{P}^1_{\mathbb{Q}}$ (therefore, the notions of cusp given in definition 1.1 and in definition 2.5 coincide for congruence subgroups). On the one hand, ∞ is the fixed point of the parabolic element $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. And we already know that, for each $s = \frac{p}{q} \in \mathbb{Q}$ (with $(p,q) = 1$), there exists $\gamma = \left(\begin{smallmatrix} p & u \\ q & v \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma(\infty) = s$; therefore, s is the only fixed point of $\gamma T \gamma^{-1}$ (which must be parabolic). On the other hand, a cusp $s \in \mathbb{R}$ satisfies $cs^2 + (a+d)s - b = 0$ for some parabolic element $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ with $a + d = \pm 2$ and $c \neq 0$. But the discriminant of this equation vanishes: this means that its solution must be rational.

If γ is an elliptic element of $\mathrm{SL}_2(\mathbb{Z})$, $|\mathrm{tr}(\gamma)|$ is an integer and is < 2. Therefore, the characteristic polynomial of γ is either $X^2 + 1$ or $X^2 \pm X + 1$ and its roots are roots of unity lying in a quadratic field extension of $\mathbb{Q}$: the only such roots of unity have order dividing 4 or 6. In conclusion, the elliptic points for $\mathrm{SL}_2(\mathbb{Z})$ are the points of $\mathbb{H}$ which are $\mathrm{SL}_2(\mathbb{Z})$–equivalent to either i or $\rho = \frac{1+i\sqrt{3}}{2}$ (by theorem 1.9).

2.2 The topology of $\Gamma \backslash \mathbb{H}^*$

In the remainder of this chapter, Γ will be a congruence subgroup (we could develop this theory for general discrete subgroups of $\mathrm{SL}_2(\mathbb{R})$, as in Shimura's book [12], but we are only interested in congruence subgroups and in this case some proofs are simpler).

Our objective is to endow $\Gamma \backslash \mathbb{H}^*$ with the structure of a compact Riemann surface. We started in the previous section classifying the points which will present a special behaviour. In this section, we define the topology of $\Gamma \backslash \mathbb{H}^*$ and study its main properties.

Definition 2.10. We extend the usual topology on $\mathbb{H}$ (as a subspace of $\mathbb{R}^2$) to a *topology on* $\mathbb{H}^*$ in the following way:

(i) a fundamental system of open neighbourhoods of ∞ is formed of the sets $N_C = \{z \in \mathbb{H} : \mathrm{Im}(z) > C\} \cup \{\infty\}$ for all $C > 0$;

(ii) if $s \in \mathbb{Q}$ is a finite cusp, there exists $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ such that $\alpha(\infty) = s$ and a fundamental system of open neighbourhoods of s is formed of the sets $\alpha(N_C)$ for all $C > 0$ ($\alpha(N_C) \setminus \{s\}$ is an open disc in $\mathbb{H}$ tangent to the real axis at s).

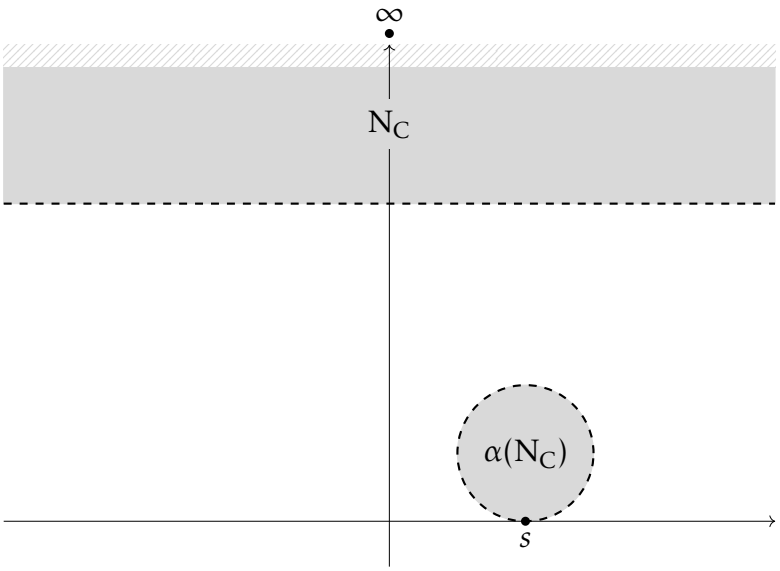


Figure 2.1: Neighbourhoods of cusps described in definition 2.10.

Proposition 2.11. $\mathbb{H}^*$ *(with the topology described in definition 2.10) is Hausdorff, connected and second-countable.*

Proof. One checks easily that $\mathbb{H}^*$ is Hausdorff (one can choose "small enough" open neighbourhoods of any two points) and second-countable (a countable base consists of open balls of rational radius centred at the points of $(\mathbb{Q} \times \mathbb{Q}) \cap \mathbb{H}$ and neighbourhoods N_C and $\alpha(N_C)$ of the cusps as described in definition 2.10 for $C \in \mathbb{Q}^+$). To prove that it is connected, we argue by contradiction. Suppose that $\mathbb{H}^* = U \cup V$ for two disjoint non-empty open sets U and V. Since $\mathbb{H}$ is connected, either $\mathbb{H} \cap U$ or $\mathbb{H} \cap V$ is empty: we may assume that $\mathbb{H} \subseteq U$ and, consequently, $V \subseteq \mathbb{P}^1_{\mathbb{Q}}$. But, in this situation, V cannot be open unless it is empty (because $\mathbb{P}^1_{\mathbb{Q}}$ contains no neighbourhoods of cusps). □

We observe that $\mathrm{SL}_2(\mathbb{R})$ is a topological group with the subspace topology (viewed as the closed subspace $\{(a,b,c,d) \in \mathbb{R}^4 : ad - bc = 1\}$ of $\mathbb{R}^4$). In particular, it is Hausdorff, locally compact and second-countable. And it is easy to see that $\mathrm{SL}_2(\mathbb{R})$ acts continuously on $\mathbb{H}$ (the map defined by this action can be seen as a rational function with non-vanishing denominators from a subset of $\mathbb{R}^4 \times \mathbb{R}^2$ to a subset of $\mathbb{R}^2$).

Lemma 2.12 (Baire's theorem). *Let* X *be a non-empty, Hausdorff and locally compact topological space. If* $\{V_n\}_{n\in\mathbb{N}}$ *is a countable family of closed subsets such that* $X = \bigcup_{n\in\mathbb{N}} V_n$*, then at least one of the sets* V_n *has an interior point.*

Proof. Suppose, for the sake of contradiction, that no V_n has an interior point. Let U_1 be any non-empty open subset of X whose closure $\overline{U_1}$ is compact. Since V_1 has no interior points, $U_1 \nsubseteq V_1$. Therefore, $U_1 \cap V_1$ is a proper relatively compact subset of the locally compact space U_1 and there exists a non-empty open subset U_2 of U_1 such that $\overline{U_2} \subseteq U_1 \setminus (U_1 \cap V_1)$. In this way, inductively, we obtain a sequence of non-empty open sets $\{U_n\}_{n\in\mathbb{N}}$ such that $\overline{U_n}$ is compact and $\overline{U_{n+1}} \subseteq U_n \setminus (U_n \cap V_n)$ for all $n \in \mathbb{N}$. The sets $\overline{U_n}$ form a decreasing sequence of non-empty compact sets and, by Cantor's intersection theorem, $\bigcap_{n\in\mathbb{N}} \overline{U_n} \neq \emptyset$. This contradicts the fact that $X = \bigcup_{n\in\mathbb{N}} V_n$. □

Proposition 2.13. *Let* G *be a topological group acting continuously and transitively on a topological space* X*. If* G *is Hausdorff, locally compact and second-countable and* X *is Hausdorff and locally compact, then the map*

$$\begin{aligned} f_x \colon G/G_x &\longrightarrow X \\ gG_x &\longmapsto gx \end{aligned}$$

is a homeomorphism for all $x \in X$*.*

Proof. f_x is clearly a bijection and is continuous by definition (because the action is continuous). We only have to show that it is open. Let U be an open subset of G and let $g \in U$. We want to prove that gx is an interior point of Ux.

The map $(h,k) \mapsto ghk : G \times G \to G$ is continuous and maps $(1,1)$ to $g \in U$. Therefore, there exists a neighbourhood V of 1, which we can take to be compact, such that $V \times V$ is mapped into U. In particular, $gVV \subseteq U$ and, after replacing V with $V \cap V^{-1}$, we can assume that $V^{-1} = V$ (where $V^{-1} = \{h^{-1} : h \in V\}$).

We can express G as the union of the interiors of the sets gV for $g \in G$. Now, we fix a countable base $\{W_n\}_{n\in\mathbb{N}}$ of G. The sets W_i contained in the interior of

some gV form a countable cover of G and we only need to take enough sets of the form gV to cover each of these W_i at least once. We obtain thus a sequence $\{g_n\}_{n\in\mathbb{N}}$ such that the interiors of the sets g_nV for $n \in \mathbb{N}$ form an open cover of G. Since, for each $n \in \mathbb{N}$, g_nV is compact, its image g_nVx is a compact subset of the Hausdorff space X and, in particular, g_nVx is closed. Now, by Baire's theorem, there exists some $n \in \mathbb{N}$ such that g_nVx has an interior point. But multiplication by g_n defines a homeomorphism between Vx and g_nVx, which means that Vx has an interior point too. That is, there exist a point $hx \in Vx$ and an open subset W of X such that $hx \in W \subseteq Vx$. But we can write

$$gx = gh^{-1} \cdot hx \in gh^{-1}W \subseteq gVVx \subseteq Ux$$

and this proves that gx is an interior point of Ux. □

Proposition 2.14. *Let* G *be a Hausdorff and locally compact topological group acting continuously and transitively on a topological space* X*. Suppose in addition that, for one (hence, for every) point* $x_0 \in X$*, the stabiliser* K *of* x_0 *in* G *is compact and the map* $gK \mapsto gx_0 : G/K \to X$ *is a homeomorphism. The following conditions on a subgroup* H *of* G *are equivalent:*

(a) *for all compact subsets* A *and* B *of* X, $\{h \in H : hA \cap B \neq \emptyset\}$ *is finite;*

(b) H *is a discrete subgroup of* G*.*

Proof. First we prove that (a) $\Longrightarrow$ (b). We consider the continuous map

$$\begin{aligned} p\colon G &\to X \\ g &\mapsto gx_0 \end{aligned}$$

and we will prove that $p^{-1}(A)$ is compact for any compact subset A of X. Consider an open cover $G = \bigcup_{i\in I} V_i$ where the sets V_i are open with compact closures $\overline{V_i}$. Observe that p is an open map because it is the composition of the projection $\pi\colon G \twoheadrightarrow G/K$ (which is open: for every open subset U of G, $\pi^{-1}(\pi(U)) = \bigcup_{k\in K} Uk$ is open) and the homeomorphism $gK \mapsto gx_0 : G/K \to X$. Since $\{p(V_i)\}_{i\in I}$ is an open cover of the compact set A, there is a finite subcover $\{p(V_j)\}_{j\in J}$ and we obtain that $p^{-1}(A) \subseteq \bigcup_{j\in J} V_jK \subseteq \bigcup_{j\in J} \overline{V_j}K$, but this is a finite union of compact sets ($\overline{V_j}K$ is the image of $\overline{V_j} \times K$ under the multiplication map). In conclusion, $p^{-1}(A)$ is a closed subset of a compact set and so is compact as well.

Let A and B be two compact subsets of X and let $h \in H$ such that $hA \cap B \neq \emptyset$. Then $p^{-1}(hA \cap B) = h \cdot p^{-1}(A) \cap p^{-1}(B) \neq \emptyset$, whence $h \in H \cap \left[p^{-1}(B) \cdot (p^{-1}(A))^{-1}\right]$.

But, since $p^{-1}(A)$ and $p^{-1}(B)$ are compact, $p^{-1}(B)\cdot(p^{-1}(A))^{-1}$ is also compact (it is the image of $p^{-1}(B)\times(p^{-1}(A))^{-1}$ under the multiplication map). In conclusion, $H\cap\left[p^{-1}(B)\cdot(p^{-1}(A))^{-1}\right]$ is the intersection of a discrete set with a compact set and it must be finite.

Now we prove that (b) $\Longrightarrow$ (a). Let V be an open neighbourhood of 1 in G whose closure $\overline{V}$ is compact. For all $x\in X$, $\{x\}$ and $\overline{V}x$ are compact and, consequently, $\{h\in H : \{hx\}\cap\overline{V}x\neq\emptyset\}$ is finite. And, since $H\cap V\subseteq\{h\in H : hx\in\overline{V}x\}$, we conclude that $H\cap V$ is also finite. This means that 1 is an isolated point of H (because G is Hausdorff) and, therefore, H is discrete. □

Proposition 2.15. *Let* G *be a Hausdorff and locally compact topological group acting continuously and transitively on a topological space* X. *Suppose in addition that, for one (hence, for every) point* $x_0\in X$, *the stabiliser* K *of* x_0 *in* G *is compact and the map* $gK\mapsto gx_0 : G/K\to X$ *is a homeomorphism (as in proposition 2.14). Let* H *be a discrete subgroup of* G.

(1) *For every* $x\in X$, $\{h\in H : hx=x\}$ *is finite.*
(2) *For each* $x\in X$, *there exists a neighbourhood* U *of* x *with the following property: if* $h\in H$ *and* $U\cap hU\neq\emptyset$, *then* $hx=x$.
(3) *For every two points* x *and* y *of* X *which are not* H*–equivalent, there exist neighbourhoods* U *of* x *and* V *of* y *such that* $hU\cap V=\emptyset$ *for all* $h\in H$.

Proof. Since H is discrete, we can use the condition (a) of proposition 2.14.

(1) Consider the map p defined by $g\mapsto gx : G\to X$. Since $\{x\}$ is compact, $p^{-1}(x)$ is also compact (see the proof of proposition 2.14). Now it is clear that $\{h\in H : hx=x\}=H\cap p^{-1}(x)$ is finite (it is the intersection of a discrete set and a compact set).
(2) Let V be any compact neighbourhood of x. Consider the (finite) subset $H'=\{h_1,\dots,h_n\}=\{h\in H : V\cap hV\neq\emptyset\}$ of H and suppose that $h_1,\dots,h_r$ are the elements of H′ which fix x. For each $i>r$, we choose neighbourhoods V_i of x and W_i of h_ix such that $V_i\cap W_i=\emptyset$ (X is Hausdorff).

$$U=V\cap\left(\bigcap_{i>r}(V_i\cap h_i^{-1}W_i)\right)$$

satisfies the required property: for $i>r$, $h_iU\subseteq W_i$ but $W_i\cap V_i=\emptyset$ and $V_i\subseteq U$.
(3) Let A and B be compact neighbourhoods of x and y, respectively. Consider the (finite) subset $\{h_1,\dots,h_n\}=\{h\in H : hA\cap B\neq\emptyset\}$ of H. We know that,

for each i, $h_i x \neq y$ (because x and y are not H–equivalent), so we can choose neighbourhoods U_i of $h_i x$ and V_i of y such that $U_i \cap V_i = \emptyset$. The neighbourhoods $U = A \cap h_1^{-1}U_1 \cap \cdots \cap h_n^{-1}U_n$ of x and $V = B \cap V_1 \cap \cdots \cap V_n$ of y satisfy the required property. □

Now we find ourselves in a position to characterise the topology of the quotient space $\Gamma \setminus \mathbb{H}$.

Theorem 2.16. *The orbit space $\Gamma \setminus \mathbb{H}$ (with the quotient topology) is Hausdorff, second-countable and connected.*

Proof. Let $\pi \colon \mathbb{H} \twoheadrightarrow \Gamma \setminus \mathbb{H}$ be the projection map. Observe that π is continuous by definition of the quotient topology and is open because $\pi^{-1}(\pi(U)) = \bigcup_{\gamma \in \Gamma} \gamma(U)$ for any open subset U of $\mathbb{H}$. Since $\Gamma \setminus \mathbb{H} = \pi(\mathbb{H})$, it is connected and second-countable (because so is $\mathbb{H}$).

Since $\mathrm{SL}_2(\mathbb{R})_i = \mathrm{SO}_2(\mathbb{R})$ is compact, the action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{H}$ satisfies all the hypotheses of propositions 2.13 to 2.15. And we know that Γ is a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$. Thus, by proposition 2.15, if Γx and Γy are two distinct points of $\Gamma \setminus \mathbb{H}$, there exist neighbourhoods (in $\mathbb{H}$) U of x and V of y such that $\gamma U \cap V = \emptyset$ for all $\gamma \in \Gamma$: then $\pi(U)$ and $\pi(V)$ are disjoint neighbourhoods of Γx and Γy, respectively. In conclusion, $\Gamma \setminus \mathbb{H}$ is Hausdorff. □

The kind of arguments which we have used can be extended to study the topology of $\Gamma \setminus \mathbb{H}^*$.

Proposition 2.17. *Let s be a cusp.*

(1) *There exists a neighbourhood U of s in $\mathbb{H}^*$ with the following property: if $\gamma \in \Gamma$ and $U \cap \gamma(U) \neq \emptyset$, then $\gamma(s) = s$.*
(2) *For every compact subset K of $\mathbb{H}$, there exists a neighbourhood V of s such that $V \cap \gamma(K) = \emptyset$ for all $\gamma \in \Gamma$.*

Proof. Recall that there exists $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ such that $\alpha(\infty) = s$. Furthermore,

$$\mathrm{SL}_2(\mathbb{Z})_\infty = \left\{ \pm \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} : h \in \mathbb{Z} \right\}$$

and, in particular, every $\beta = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \setminus \mathrm{SL}_2(\mathbb{Z})_\infty$ satisfies that $|c| \geq 1$ and so

$$\mathrm{Im}(z) \cdot \mathrm{Im}(\beta(z)) = \frac{\mathrm{Im}(z)^2}{|cz + d|^2} \leq 1$$

because $\mathrm{Im}(z) \leq |c|\mathrm{Im}(z) \leq |cz + d|$.

(1) Let $N_1 = \{z \in \mathbb{H} : \operatorname{Im}(z) > 1\} \cup \{\infty\}$. I claim that $U = \alpha(N_1)$ has the required property. If $U \cap \gamma(U) \neq \emptyset$, then $N_1 \cap (\alpha^{-1}\gamma\alpha)(U) \neq \emptyset$. But, for all $z \in N_1 \setminus \{\infty\}$, $\operatorname{Im}(z) > 1$ implies that $\operatorname{Im}((\alpha^{-1}\gamma\alpha)(z)) < 1$; thus, $(\alpha^{-1}\gamma\alpha)(z) \notin N_1$ and also $\gamma(\alpha(z)) \notin U$.

(2) Since K is compact, $0 < A < \operatorname{Im}(\alpha^{-1}(z)) < B$ for all $z \in K$ for some constants A and B. Let $C = \max\{B, \frac{1}{A}\}$ and let $N_C = \{z \in \mathbb{H} : \operatorname{Im}(z) > C\} \cup \{\infty\}$. We can choose $V = \alpha(N_C)$. Let $z \in K$.

- If $\gamma \in \Gamma_s = \alpha\Gamma_\infty\alpha^{-1}$, $\alpha^{-1}\gamma\alpha \in \Gamma_\infty$ and thus

$$\operatorname{Im}((\alpha^{-1}\gamma)(z)) = \operatorname{Im}((\alpha^{-1}\gamma\alpha)(\alpha^{-1}(z))) = \operatorname{Im}(\alpha^{-1}(z)) < B \leq C.$$

  Consequently, $(\alpha^{-1}\gamma)(z) \notin N_C$ and $\gamma(z) \notin V$.
- If $\gamma \in \Gamma \setminus \Gamma_s$, we know that $\operatorname{Im}((\alpha^{-1}\gamma\alpha)(\alpha^{-1}(z))) \cdot \operatorname{Im}(\alpha^{-1}(z)) < 1$. Therefore, $\operatorname{Im}((\alpha^{-1}\gamma)(z)) < \frac{1}{A} \leq C$, which implies that $(\alpha^{-1}\gamma)(z) \notin N_C$ and $\gamma(z) \notin V$. □

Theorem 2.18. *The quotient space $\Gamma \setminus \mathbb{H}^*$ is second-countable, Hausdorff, connected and compact.*

Proof. Observe that the quotient map $\pi : \mathbb{H}^* \twoheadrightarrow \Gamma \setminus \mathbb{H}^*$ is continuous and open, as in the proof of theorem 2.16. Since $\Gamma \setminus \mathbb{H}^* = \pi(\mathbb{H}^*)$, $\Gamma \setminus \mathbb{H}^*$ is second-countable and connected (because so is $\mathbb{H}^*$).

Let us prove that $\Gamma \setminus \mathbb{H}^*$ is Hausdorff. Theorem 2.16 asserts that $\Gamma \setminus \mathbb{H}$ is Hausdorff, so we have to prove that an equivalence class of cusps can be separated from an equivalence class of points in $\mathbb{H}$ and also from another equivalence class of cusps.

If $z \in \mathbb{H}$ and $s \in \mathbb{P}^1_{\mathbb{Q}}$, let K be a compact neighbourhood of z in $\mathbb{H}$ and there exists a neighbourhood U of s such that $U \cap \gamma(K) = \emptyset$ for all $\gamma \in \Gamma$, by proposition 2.17. In this situation, $\pi(U)$ and $\pi(K)$ are disjoint neighbourhoods of Γs and Γz, respectively.

Let s and t be two cusps in different orbits under the action of Γ and let $\alpha, \beta \in \mathrm{SL}_2(\mathbb{Z})$ such that $\alpha(\infty) = s$ and $\beta(\infty) = t$. We take $U = \alpha(N_2)$ and $V = \beta(N_2)$, where $N_2 = \{z \in \mathbb{H} : \operatorname{Im}(z) > 2\} \cup \{\infty\}$, and claim that $\pi(U)$ and $\pi(V)$ are disjoint neighbourhoods of Γs and Γt, respectively. Suppose, for the sake of contradiction, that $(\gamma\alpha)(z) = \beta(w)$ for some $z, w \in N_2 \setminus \{\infty\}$ and some $\gamma \in \Gamma$. Then $\beta^{-1}\gamma\alpha$ maps z to w. Let F be the fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$ described in theorem 1.9. Observe that $N_2 \setminus \{\infty\}$ is tessellated by the integer translates of F. Furthermore,

$N_2 \setminus \{\infty\}$ contains no elliptic points (the only points of $\mathbb{H}$ whose stabilisers under the action of $SL_2(\mathbb{Z})$ are non-trivial). Therefore, $\beta^{-1}\gamma\alpha$ must be of the form $\pm\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right)$ for some $h \in \mathbb{Z}$. In particular, $(\beta^{-1}\gamma\alpha)(\infty) = \infty$ and so $\gamma(s) = t$, contradicting thus our choice of s and t.

Finally, we prove that $\Gamma \setminus \mathbb{H}^*$ is compact. To this aim, we have to prove that $\widetilde{F} = F \cup \{\infty\}$ is a compact subset of $\mathbb{H}^*$. Let $\{U_i\}_{i\in I}$ be an open cover of $\widetilde{F}$. Since $\infty \in \widetilde{F}$, there is some $i_0 \in I$ such that $\infty \in U_{i_0}$. And, by definition 2.10, there exists some $C > 0$ such that $N_C = \{z \in \mathbb{H} : \operatorname{Im}(z) > C\} \cup \{\infty\} \subseteq U_{i_0}$. Now it is clear that $\widetilde{F} \setminus N_C$ is compact (since it is a closed and bounded subset of $\mathbb{R}^2$) and, therefore, it has a finite subcover $\{U_{i_1}, \dots, U_{i_m}\}$. As a result, $\widetilde{F} \subseteq U_{i_0} \cup U_{i_1} \cup \cdots \cup U_{i_m}$.

Let $\alpha_1, \dots, \alpha_n$ be a set of representatives of the left cosets of Γ in $SL_2(\mathbb{Z})$. By proposition 1.10, we know that $D = \bigcup_{j=1}^{n} \alpha_j^{-1}\widetilde{F}$ contains at least one representative of each Γ–orbit. And, since α_j^{-1} is a homeomorphism for each j, D is compact. In conclusion, $\Gamma \setminus \mathbb{H}^* = \pi(D)$ and this is compact. □

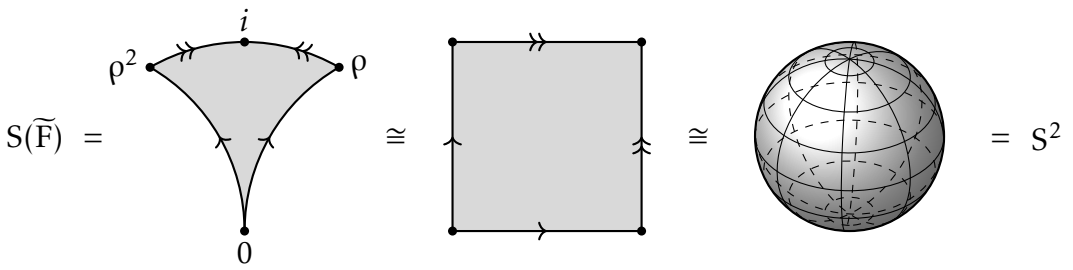


Figure 2.2: The topology of $SL_2(\mathbb{Z}) \setminus \mathbb{H}^*$.

Example 2.19. The fundamental domain F for $SL_2(\mathbb{Z})$ described in theorem 1.9 may be visualised as a triangle with a vertex removed in the Riemann sphere. That is why we need to add a point (the cusp ∞) to compactify it: let $\widetilde{F} = F \cup \{\infty\}$. As we saw in the last part of the proof of theorem 2.18, $SL_2(\mathbb{Z}) \setminus \mathbb{H}^* = \pi(\widetilde{F})$, where $\pi : \mathbb{H}^* \twoheadrightarrow SL_2(\mathbb{Z}) \setminus \mathbb{H}^*$ is the quotient map. Therefore, $SL_2(\mathbb{Z}) \setminus \mathbb{H}^*$ is homeomorphic to the triangle $\widetilde{F}$ with its sides identified according to π. Figure 2.2 illustrates that $SL_2(\mathbb{Z}) \setminus \mathbb{H}^*$ is homeomorphic to a sphere using a translate of $\widetilde{F}$. Thus, $\Gamma \setminus \mathbb{H}^*$ is also homeomorphic to a polygon with sides identified.

2.3 The complex structure on $\Gamma \setminus \mathbb{H}^*$

A Riemann surface is a Hausdorff and connected topological space endowed with a complex structure. In this section, we describe explicitly an atlas of coordinate

charts on $\Gamma \backslash \mathbb{H}^*$ whose changes of coordinates are analytic. The same charts also give $\Gamma \backslash \mathbb{H}$ a complex structure (restricting the domains and codomains when necessary).

For any $r \in \mathbb{R}^+$, let $\mathrm{D}_r = \{z \in \mathbb{C} : |z| < r\}$. We recall some results which are going to be useful in order to define a set of charts.

Lemma 2.20 (Schwarz). *Let $f : \mathrm{D}_1 \to \mathrm{D}_1$ be a holomorphic map such that $f(0) = 0$. Then $|f(z)| \leq |z|$ for all $z \in \mathrm{D}_1$ and $f'(0) \leq 1$. Moreover, if $|f(z_0)| = |z_0|$ for some $z_0 \in \mathrm{D}_1 \setminus \{0\}$ or $|f'(0)| = 1$, $f(z) = \lambda z$ for some $\lambda \in \mathbb{C}$ with $|\lambda| = 1$.*

Proof. Consider $g\colon \mathrm{D}_1 \to \mathbb{C}$ defined by

$$g(z) = \begin{cases} \frac{f(z)}{z} & \text{if } z \neq 0, \\ f'(0) & \text{if } z = 0, \end{cases}$$

which is holomorphic because $f(0) = 0$. For all $0 < r < 1$, the maximum modulus principle asserts that there exists $z_r \in \partial \mathrm{D}_r$ such that

$$|g(z)| \leq |g(z_r)| = \frac{|f(z_r)|}{|z_r|} < \frac{1}{r} \quad \forall z \in \overline{\mathrm{D}_r}.$$

Thus, taking the limit as $r \to 1^-$, we obtain that $|g(z)| \leq 1$ for all $z \in \mathrm{D}_1$. Moreover, if $|g(z)| = 1$ for some $z \in \mathrm{D}_1$, $g(z)$ must be equal to a constant λ (again by the maximum modulus principle) with $|\lambda| = 1$. □

Lemma 2.21. *The analytic automorphisms of D_1 fixing 0 are the maps of the form $z \mapsto \lambda z$ with $|\lambda| = 1$.*

Proof. If $f : \mathrm{D}_1 \to \mathrm{D}_1$ is an analytic automorphism of D_1 with $f(0) = 0$, Schwarz's lemma implies that $|f(z)| \leq |z|$ and $|f^{-1}(z)| \leq |z|$ for all $z \in \mathrm{D}_1$. Hence, $|f(z)| = |z|$ and, consequently, $f(z) = \lambda z$ for some $\lambda \in \partial \mathrm{D}_1$. □

Proposition 2.22. *For all $v \in \mathbb{H}^*$, there exists an open neighbourhood U of v such that $\Gamma_v = \{\gamma \in \Gamma : \gamma(\mathrm{U}) \cap \mathrm{U} \neq \emptyset\}$.*

Proof. It is immediate from propositions 2.15 and 2.17. □

Let $\pi\colon \mathbb{H}^* \twoheadrightarrow \Gamma \backslash \mathbb{H}^*$ be the quotient map. Recall that π is continuous (by definition of the quotient topology) and open (because $\pi^{-1}(\pi(\mathrm{U})) = \bigcup_{\gamma \in \Gamma} \gamma(\mathrm{U})$ for every open set U and each $\gamma \in \Gamma$ is a diffeomorphism). Let $p \in \Gamma \backslash \mathbb{H}^*$ and

consider $v \in \mathbb{H}^*$ such that $\pi(v) = p$. Consider an open neighbourhood U of v with the property that $\Gamma_v = \{\gamma \in \Gamma : \gamma(U) \cap U \neq \emptyset\}$ (proposition 2.22 asserts the existence of such U). Assume further that U is a domain (i.e., open and connected), up to replacing it with some fundamental open neighbourhood of v contained in it. Now we write $\overline{\Gamma_v} = (\{\pm 1\} \cdot \Gamma_v)/\{\pm 1\}$ (this is the associated group of transformations of $\mathbb{H}^*$). Observe that $\pi(U) = \Gamma \backslash U$ is an open neighbourhood of p. Next, we are going to define a chart of $X(\Gamma)$ given by a map $\varphi\colon \Gamma \backslash U \to V$ for some open subset V of $\mathbb{C}$. To do so, we must distinguish three cases.

(1) If v is neither an elliptic point nor a cusp, $\overline{\Gamma_v}$ is trivial. In particular, no two distinct points of U are Γ–equivalent, so the restricted map $\pi|_U\colon U \to \Gamma \backslash U$ is a homeomorphism. Let $\varphi\colon \Gamma \backslash U \to U$ be its inverse. We can take $(\Gamma \backslash U, \varphi)$ as a chart at p.

(2) If v is an elliptic point, we know that $\overline{\Gamma_v}$ is a cyclic group of order d by proposition 2.6 (in fact, we even know that d is 2 if v is $\mathrm{SL}_2(\mathbb{Z})$–equivalent to i and 3 if it is $\mathrm{SL}_2(\mathbb{Z})$–equivalent to ρ, by theorem 1.9).
The linear fractional transformation corresponding to $\delta = \left(\begin{smallmatrix} 1 & -v \\ 1 & -\overline{v} \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{C})$ defines an analytic isomorphism from $\mathbb{H}$ to D_1 which maps v to 0. Thus, $\delta\overline{\Gamma_v}\delta^{-1}$ is a cyclic subgroup of automorphisms of D_1 of order d: lemma 2.21 implies that its elements are of the form $z \mapsto \zeta_d z$ where ζ_d is a d–th root of unity.

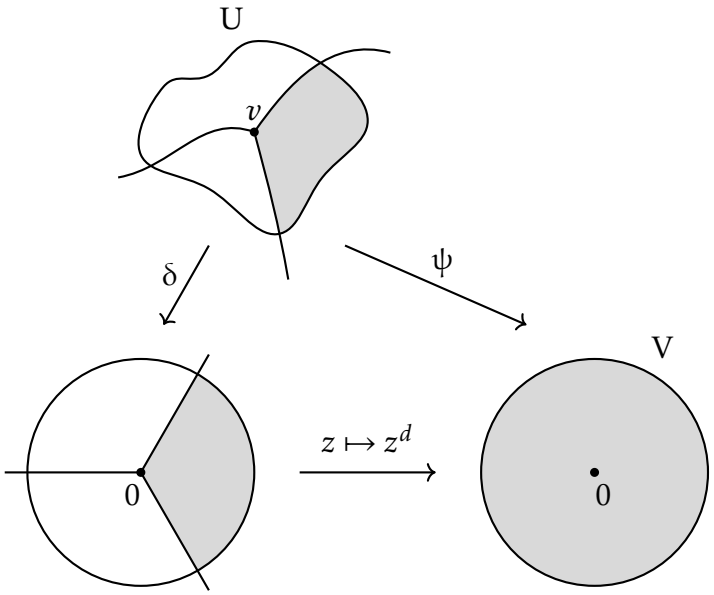


Figure 2.3: Definition of local coordinates at an elliptic point.

Let ψ denote the holomorphic map $z \mapsto \delta(z)^d : U \to \mathbb{C}$, which is open by

the open mapping theorem, and let $V = \psi(U)$. Figure 2.3 illustrates the situation for $d = 3$: the neighbourhood U is homeomorphic to a disc whose points are grouped in $\Gamma_v \setminus U$ in orbits of d elements; the map δ "straightens" the disc and the map $z \mapsto z^d$ folds it identifying the points in the same way as π, as we shall see.

For any two points $z, z' \in U$, we have that $\pi(z) = \pi(z')$ if and only if $z' = \gamma(z)$ or, equivalently, $\delta(z') = (\delta\gamma\delta^{-1})(\delta(z)) = \zeta_d\delta(z)$ (where ζ_d is a d–th root of unity) for some $\gamma \in \Gamma_v$. That is, $\pi(z') = \pi(z)$ if and only if $\psi(z') = \psi(z)$. In conclusion, ψ descends to the quotient (universal property of the quotient topology): there exists a continuous bijection φ making the diagram

$$\begin{array}{ccc} U & \xrightarrow{\psi} & V \\ {\scriptstyle\pi}\downarrow & \nearrow_{\varphi} & \\ \Gamma \setminus U & & \end{array}$$

commutative. Furthermore, φ is open because ψ is. Thus, we take $(\Gamma \setminus U, \varphi)$ as a chart at p.

(3) If v is a cusp, there exists some $\delta \in \mathrm{SL}_2(\mathbb{Z})$ such that $\delta(v) = \infty$. Then

$$\{\pm 1\} \cdot \delta\Gamma_v\delta^{-1} = \left\{ \pm\begin{pmatrix} 1 & mh \\ 0 & 1 \end{pmatrix} : m \in \mathbb{Z} \right\}$$

for some positive integer h because this is a subgroup of

$$\mathrm{SL}_2(\mathbb{Z})_\infty = \left\{ \pm\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} : m \in \mathbb{Z} \right\}.$$

In particular, $h = [\mathrm{SL}_2(\mathbb{Z})_\infty : (\{\pm 1\} \cdot \delta\Gamma_v\delta^{-1})]$. In this case, the order of the stabiliser Γ_v is infinite; that is why we need to "measure" its size looking at this index. Let ψ denote the holomorphic map $z \mapsto e^{2\pi i\delta(z)/h} : U \to \mathbb{C}$, which is open by the open mapping theorem, and let $V = \psi(U)$. Figure 2.4 illustrates the situation in the case in which U is a fundamental neighbourhood of v: the points of U are grouped in $\Gamma_v \setminus U$ in orbits containing infinitely many elements; the map δ translates the disc to a half plane (in which equivalent points differ by a horizontal offset) and the map $z \mapsto e^{2\pi iz/h}$ folds it identifying the points in the same way as π, as we shall see.

For any two points $z, z' \in U$, we have that $\pi(z) = \pi(z')$ if and only if $z' = \gamma(z)$ or, equivalently, $\delta(z') = (\delta\gamma\delta^{-1})(\delta(z)) = \delta(z) + mh$ (where $m \in \mathbb{Z}$) for some $\gamma \in \Gamma_v$. That is to say, $\pi(z') = \pi(z)$ if and only if $\psi(z') = \psi(z)$. In conclusion,

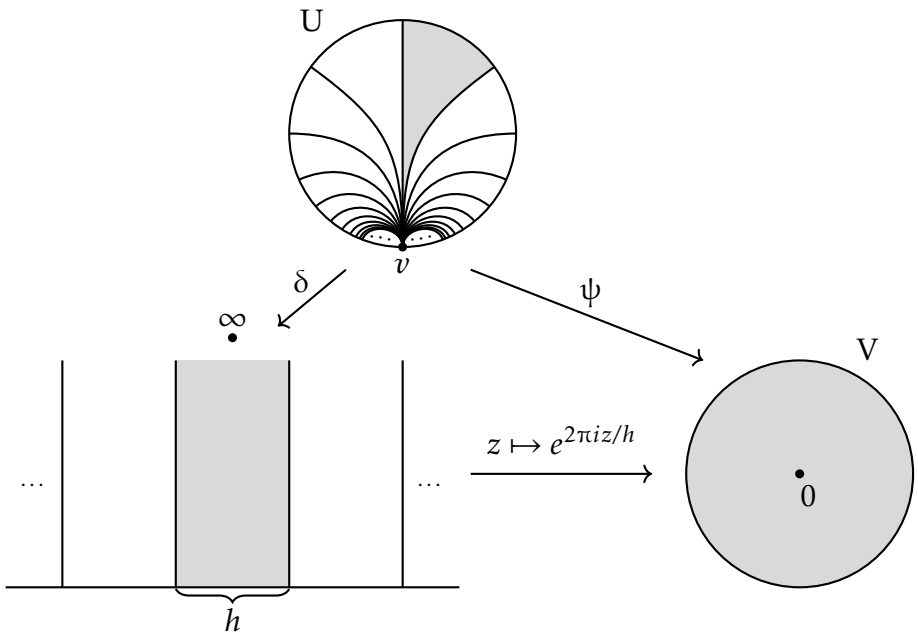
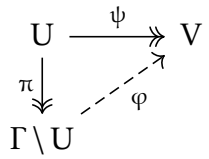


Figure 2.4: Definition of local coordinates at a cusp.

ψ descends to the quotient (universal property of the quotient topology): there exists a continuous bijection φ making the diagram



commutative. Furthermore, φ is open because ψ is. Thus, we take $(\Gamma \backslash U, \varphi)$ as a chart at p.

Theorem 2.23. *The charts described above endow $\Gamma \backslash \mathbb{H}^*$ (and also $\Gamma \backslash \mathbb{H}$ by restriction) with a complex structure.*

Proof. Since the domains of the given charts form an open cover of $\Gamma \backslash \mathbb{H}^*$, we only have to check that the transition maps are holomorphic.

Let $v \in \mathbb{H}^*$ and consider an open neighbourhood U of v with the property that $\Gamma_v = \{\gamma \in \Gamma : \gamma(U) \cap U \neq \emptyset\}$. Observe that, by definition, U contains no elliptic points or cusps apart from possibly v. Now consider two distinct points p_1 and p_2 of $\Gamma \backslash \mathbb{H}^*$. Let $v_1 \in \pi^{-1}(p_1)$ and $v_2 \in \pi^{-1}(p_2)$. Let U_1 and U_2 be two domains such that $v_i \in U_i$ and $\Gamma_{v_i} = \{\gamma \in \Gamma : \gamma(U_i) \cap U_i \neq \emptyset\}$ for $i = 1$ and 2. Consider the two corresponding charts $\varphi_1 : \Gamma \backslash U_1 \to V_1$ and $\varphi_2 : \Gamma \backslash U_2 \to V_2$ (as defined above). Write $W = \Gamma \backslash U_1 \cap \Gamma \backslash U_2$. Our goal is to prove that the transition map

$\phi = (\varphi_1 \circ \varphi_2^{-1})\big|_{\varphi_2(W)} \colon \varphi_2(W) \to \varphi_1(W)$ is holomorphic. To do so, we are going to prove that ϕ is holomorphic at every point of $\varphi_2(W)$. Let $x \in W$ and write $x = \pi(z_1) = \pi(z_2)$ with $z_1 \in U_1$ and $z_2 \in U_2$. Then $z_1 = \gamma(z_2)$ for some $\gamma \in \Gamma$ and $U = \gamma^{-1}(U_1) \cap U_2$ is an open neighbourhood of z_2 with $\pi(U) \subseteq W$. Now we distinguish several cases.

- If v_1 and v_2 are not elliptic points or cusps, $V_1 = U_1$ and $V_2 = U_2$. Each of these two domains contains exactly one representative of every element of W: ϕ maps the representative in U_2 to the representative in U_1. Therefore, $\phi(z) = \gamma(z)$ for all $z \in U$ and, in particular, ϕ is holomorphic in U.
- If v_1 is an elliptic point (and so v_2 is neither an elliptic point nor a cusp), φ_2^{-1} coincides with π. In this case,

$$\phi(z) = \varphi_1(\pi(z)) = \varphi_1(\pi(\gamma(z))) = \psi_1(\gamma(z)) = (\delta_1(\gamma(z)))^{d_1}$$

  for all $z \in U$, where $\delta_1 = \left(\begin{smallmatrix} 1 & -v_1 \\ 1 & -\overline{v_1} \end{smallmatrix}\right)$ and $d_1 = |\overline{\Gamma_v}|$. Thus, ϕ is holomorphic in U.
- If v_2 is an elliptic point (and so v_1 is neither an elliptic point nor a cusp), ϕ is holomorphic because it is a bijection between open subsets of $\mathbb{C}$ and its inverse ϕ^{-1} is holomorphic (by the previous case).
- If v_1 is a cusp (and so v_2 is neither an elliptic point nor a cusp), φ_2^{-1} coincides with π. Thus, for all $z \in U$,

$$\phi(z) = \varphi_1(\pi(z)) = \varphi_1(\pi(\gamma(z))) = \psi_1(\gamma(z)) = e^{2\pi i \delta_1(\gamma(z))/h_1},$$

  where $\delta_1 \in SL_2(\mathbb{Z})$ with $\delta_1(v_1) = \infty$ and $h_1 = [SL_2(\mathbb{Z})_\infty : (\{\pm 1\} \cdot \delta_1 \Gamma_{v_1} \delta_1^{-1})]$. Hence, ϕ is holomorphic in U.
- If v_2 is a cusp (and so v_1 is neither an elliptic point nor a cusp), ϕ is holomorphic because it is a bijection between open subsets of $\mathbb{C}$ and its inverse ϕ^{-1} is holomorphic (by the previous case).
- In all the other cases, $W = \emptyset$. □

Definition 2.24. The *modular curve* $Y(\Gamma)$ is the Riemann surface $\Gamma \backslash \mathbb{H}$. The *compactified modular curve* $X(\Gamma)$ is the Riemann surface $\Gamma \backslash \mathbb{H}^*$.

For all $N \in \mathbb{N}$, we abbreviate $Y(\Gamma(N))$ to $Y(N)$, $X(\Gamma(N))$ to $X(N)$, $Y(\Gamma_0(N))$ to $Y_0(N)$, $X(\Gamma_0(N))$ to $X_0(N)$, $Y(\Gamma_1(N))$ to $Y_1(N)$ and $X(\Gamma_1(N))$ to $X_1(N)$.

Observe that the charts at the cusps resemble the changes of variables used to define the conditions at the cusps in section 1.2. Later we shall see that modular forms for Γ of weight $2k$ can be seen as k–fold differential forms on $X(\Gamma)$.

2.4 Dimension formulae

Since $X(\Gamma)$ is a compact Riemann surface, we can deduce important facts about $X(\Gamma)$ and even about modular forms for Γ using results from the theory of Riemann surfaces.

We have studied modular forms for $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$ in chapter 1. We are going to use these results in order to study modular forms for Γ. One checks easily that the natural projection

$$\begin{aligned} \phi\colon X(\Gamma) &\longrightarrow X(1) \\ \Gamma v &\longmapsto \Gamma(1)v \end{aligned}$$

is holomorphic (the arguments are analogous to those of the proof of theorem 2.23) and, in fact, it is a covering of degree $[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}]$. (Recall that the linear fractional transformations do not depend on the sign of the corresponding matrix: as in chapter 1, a bar denotes the image of elements and subgroups of $\mathrm{SL}_2(\mathbb{R})$ in $\mathrm{PSL}_2(\mathbb{R})$.) Indeed, if $\mathrm{PSL}_2(\mathbb{Z}) = \bigsqcup_{j=1}^{m} \overline{\Gamma}\overline{\alpha_j}$, then $\phi^{-1}(\Gamma(1)v) = \{\Gamma\alpha_j(v)\}_{j=1}^{m}$ for all $v \in \mathbb{H}^*$ (and the elements $\Gamma\alpha_j(v)$ are all distinct if v is neither an elliptic element nor a cusp). Hence, we can apply the Riemann–Hurwitz formula to ϕ.

Theorem 2.25. *Let $m = [\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}]$. Let v_2 be the number of inequivalent elliptic points for Γ which are $\mathrm{SL}_2(\mathbb{Z})$–equivalent to i, let v_3 be the number of inequivalent elliptic points for Γ which are $\mathrm{SL}_2(\mathbb{Z})$–equivalent to $\rho = e^{\pi i/3}$ and let v_∞ be the number of inequivalent cusps for Γ. The genus of $X(\Gamma)$ is*

$$g = 1 + \frac{m}{12} - \frac{v_2}{4} - \frac{v_3}{3} - \frac{v_\infty}{2}.$$

Proof. Since $X(1)$ has genus 0, the Riemann–Hurwitz formula states that

$$2g - 2 = m(2 \cdot 0 - 2) + \sum_{p \in X(\Gamma)} (e_p(\phi) - 1),$$

where $e_p(\phi)$ is the ramification index of ϕ at p. That is,

$$g = 1 - m + \sum_{p \in X(\Gamma)} \frac{e_p(\phi) - 1}{2}.$$

Let $\pi\colon \mathbb{H}^* \to X(\Gamma)$ and $\pi_1\colon \mathbb{H}^* \to X(1)$ be the quotient maps. Since the ramific-

ation indices are multiplicative and $\pi_1 = \phi \circ \pi$, we have that

$$e_z(\pi_1) = e_z(\pi) \cdot e_{\pi(z)}(\phi)$$

for all $z \in \mathbb{H}^*$. If z is a cusp, this formula is not useful because $e_z(\pi_1) = e_z(\pi) = \infty$; assume, thus, that $z \in \mathbb{H}$. We know that there exists some open neighbourhood U of z such that $\mathrm{SL}_2(\mathbb{Z})_z = \{\alpha \in \mathrm{SL}_2(\mathbb{Z}) : \alpha(U) \cap U \neq \emptyset\}$. Thus, $e_z(\pi_1) = |\mathrm{PSL}_2(\mathbb{Z})_z|$. Similarly, $e_z(\pi) = |\overline{\Gamma}_z|$.

If z is not an elliptic point for $\mathrm{SL}_2(\mathbb{Z})$, $e_z(\pi_1) = 1$ and so $e_z(\pi) = e_{\pi(z)}(\phi) = 1$. Therefore, the points of the form $p = \pi(z)$ for such z do not contribute to the sum $\sum(e_p(\phi) - 1)$.

If z is $\mathrm{SL}_2(\mathbb{Z})$–equivalent to i, $e_z(\pi_1) = 2$. Therefore, either $e_z(\pi) = 2$ and $e_{\pi(z)}(\phi) = 1$ or $e_z(\pi) = 1$ and $e_{\pi(z)}(\phi) = 2$. In the first case, z is an elliptic point for Γ and ϕ is unramified at $\pi(z)$; in the second case, z is not an elliptic point for Γ and the ramification index of ϕ at $\pi(z)$ is 2. Hence, there are exactly $\frac{m-v_2}{2}$ points of the form $p = \pi(z)$ such that $e_p(\phi) = 2$.

If z is $\mathrm{SL}_2(\mathbb{Z})$–equivalent to ρ, $e_z(\pi_1) = 3$. Therefore, either $e_z(\pi) = 3$ and $e_{\pi(z)}(\phi) = 1$ or $e_z(\pi) = 1$ and $e_{\pi(z)}(\phi) = 3$. In the first case, z is an elliptic point for Γ and ϕ is unramified at $\pi(z)$; in the second case, z is not an elliptic point for Γ and the ramification index of ϕ at $\pi(z)$ is 3. Hence, there are exactly $\frac{m-v_3}{3}$ points of the form $p = \pi(z)$ such that $e_p(\phi) = 3$.

Finally, there are exactly v_∞ inequivalent cusps for Γ, and these correspond to the elements of $\phi^{-1}(\mathrm{SL}_2(\mathbb{Z})\infty)$. Therefore, $\sum_p e_p(\phi) = m$, where the sum is over the v_∞ points p of $X(\Gamma) \setminus Y(\Gamma)$.

In conclusion,

$$g = 1 - m + \frac{1}{2}\left[0 + 1\frac{m-v_2}{2} + 2\frac{m-v_3}{3} + (m - v_\infty)\right] = 1 + \frac{m}{12} - \frac{v_2}{4} - \frac{v_3}{3} - \frac{v_\infty}{2}. \qquad \square$$

Next we relate (meromorphic) modular forms for Γ with the modular curve $X(\Gamma)$ as a Riemann surface. Consider the restriction $\widetilde{\pi} = \pi|_{\mathbb{H}}$ of the quotient map $\pi\colon \mathbb{H}^* \to X(\Gamma)$. Observe that $\mathbb{H}$ is an open subset of $\mathbb{C}$ and $\widetilde{\pi}$ is holomorphic: its expressions in local coordinates are of the form $z \mapsto \delta(z)^a$ for some $\delta \in \mathrm{GL}_2(\mathbb{C})$ with $\delta(z) \neq \infty$ and some $a \in \mathbb{N}$.

Proposition 2.26. *The field of modular functions for* Γ *and the field of meromorphic functions on* $X(\Gamma)$ *are isomorphic.*

Proof. If F is a meromorphic function on $X(\Gamma)$, its pull-back $f = \widetilde{\pi}^*(F) = F \circ \widetilde{\pi}$ is also meromorphic in $\mathbb{H}$. And, by definition, it is clear that $f \circ \gamma = f$ for all $\gamma \in \Gamma$. Finally, let s be a cusp and let $(\Gamma \setminus U, \varphi)$ be the chart of $X(\Gamma)$ at $\pi(s)$ defined in section 2.3. We can assume, up to replacing s with $\gamma(s)$ for some $\gamma \in \Gamma$, that $s \in U$. The local expression $\widehat{F}(q) = (F \circ \varphi^{-1})(q)$ is meromorphic and, in a neighbourhood of s, $f(z) = \widehat{F}(e^{2\pi i \delta(z)/h})$ for some $\delta \in \mathrm{SL}_2(\mathbb{Z})$ such that $\delta(s) = \infty$ and some $h \in \mathbb{N}$; that is, f is meromorphic at s. All in all, f is a modular function for Γ.

Conversely, suppose that f is a modular function for Γ. Since $f \circ \gamma = f$ for all $\gamma \in \Gamma$, f induces a function F on $X(\Gamma)$ such that $F \circ \pi = f$ (this is well-defined). Consider a chart $(\Gamma \setminus U, \varphi)$ of $X(\Gamma)$ at a point p as defined in section 2.3 and let $v \in U$ such that $\pi(v) = p$. We must check that F is meromorphic at p.

- If v is neither an elliptic point nor a cusp, $F \circ \varphi^{-1} = f|_U$ (because φ is a local inverse of π) and F is clearly meromorphic at p.
- If v is an elliptic point, the local expression of F at p is $\widehat{F}(\tau) = f(\delta^{-1}(\tau^{1/d}))$ where $\delta = \left(\begin{smallmatrix} 1 & -v \\ 1 & -\overline{v} \end{smallmatrix}\right)$ and $d = |\overline{\Gamma_v}|$. But, since $f \circ \delta^{-1}$ is meromorphic and satisfies that $(f \circ \delta^{-1})(\zeta_d z) = (f \circ \delta^{-1})(z)$ for any d–th root of unity ζ_d, we deduce that $f \circ \delta^{-1}$ has a Laurent series expansion containing only powers of z^d. Therefore, $\widehat{F}(\tau)$ is meromorphic.
- If v is a cusp, then $(\varphi \circ \pi)(z) = e^{2\pi i \delta(z)/h}$ for all $z \in U$, with $\delta \in \mathrm{SL}_2(\mathbb{Z})$ such that $\delta(v) = \infty$ and $h = [\mathrm{SL}_2(\mathbb{Z})_\infty : \{\pm 1\} \cdot \delta \Gamma_v \delta^{-1}]$. Since f is meromorphic at v, we have that $(f \circ \delta^{-1})(z) = \widehat{f_v}(e^{2\pi i z/h})$, where $\widehat{f_v}$ is meromorphic at 0. In turn, the local expression of F at p is $\widehat{F}(q) = f(z) = (f \circ \delta^{-1})(\delta(z)) = \widehat{f_v}(q)$, where $q = e^{2\pi i \delta(z)/h}$. Therefore, $\widehat{F}(q)$ is meromorphic at $\varphi(p)$.

In conclusion, F is meromorphic.

Furthermore, this correspondence preserves sums and products. □

If M is a Riemann surface, it is also a smooth manifold of (real) dimension 2. Hence, for every point $p \in M$, we can consider the cotangent space (i.e., the dual space of the tangent space) at p, $T_p^* M$, which is a real vector space with a basis $\{dx, dy\}$ (where x and y are the local coordinates at p). By extension of scalars, we obtain the complex vector space $T_p^* M \otimes_{\mathbb{R}} \mathbb{C}$. The complex structure given by multiplication of the local coordinates by i induces a decomposition of this space as the direct sum of the eigenspaces generated by $dz = dx + i\,dy$ and by $d\overline{z} = dx - i\,dy$ (we write $z = x + iy$ and $\overline{z} = x - iy$).

A differential form ω on M is a map assigning to each point $p \in M$ (except for possibly a discrete set of points) an element of $T_p^* M \otimes_{\mathbb{R}} \mathbb{C}$. Hence, ω can be expressed in local coordinates as $f(z)\,dz + g(z)\,d\overline{z}$. We say that ω is meromorphic

(resp. holomorphic) if $g(z) = 0$ and $f(z)$ is meromorphic (resp. holomorphic).

Let $k \in \mathbb{N}$. A k–fold differential form ω on M is a map assigning to each point $p \in M$ (except for possibly a discrete set of points) an element of the tensor space $(T_p^* M \otimes_{\mathbb{R}} \mathbb{C}) \otimes_{\mathbb{C}} \overset{(k)}{\cdots} \otimes_{\mathbb{C}} (T_p^* M \otimes_{\mathbb{R}} \mathbb{C})$. If ω is meromorphic (resp. holomorphic), it can be expressed in local coordinates as $f(z)(dz)^k$, where $f(z)$ is a meromorphic (resp. holomorphic) function.

Proposition 2.27. *Let k be a positive integer. There is an isomorphism between the space of meromorphic modular forms for Γ of weight $2k$ and the space of meromorphic k–fold differential forms on $X(\Gamma)$.*

Proof. Let ω be a meromorphic k–fold differential form on $X(\Gamma)$. Its pull-back $\widetilde{\pi}^*(\omega)$ is a meromorphic k–fold differential form on $\mathbb{H}$ which can be written as $f(z)(dz)^k$. By definition, $\widetilde{\pi}^*(\omega)$ is invariant under Γ; that is,

$$f(z)(dz)^k = f(\gamma(z))(d\gamma(z))^k = f(\gamma(z))\left(\frac{d}{dz}\gamma(z)\right)^k (dz)^k = f(\gamma(z))(cz+d)^{-2k}(dz)^k$$

for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. Hence, f is weakly modular for Γ of weight $2k$. Finally, let s be a cusp and let $(\Gamma \setminus U, \varphi)$ be the chart of $X(\Gamma)$ at $\pi(s)$ defined in section 2.3. Assume, up to replacing s with $\gamma(s)$ for some $\gamma \in \Gamma$, that $s \in U$. The local expression of ω with respect to this chart is $F(q)(dq)^k$, where F is a meromorphic function. And recall that, in a neighbourhood of s, $\varphi(\pi(z)) = e^{2\pi i \delta(z)/h}$ for some $\delta \in SL_2(\mathbb{Z})$ such that $\delta(s) = \infty$ and some $h \in \mathbb{N}$. Therefore, for all z in a neighbourhood of s,

$$\begin{aligned} f(z)(dz)^k &= F(e^{2\pi i\delta(z)/h})(de^{2\pi i\delta(z)/h})^k = F(e^{2\pi i\delta(z)/h})\left(\frac{2\pi i}{h}e^{2\pi i\delta(z)/h}\delta'(z)\,dz\right)^k \\ &= F(e^{2\pi i\delta(z)/h})\left(\frac{2\pi i}{h}e^{2\pi i\delta(z)/h}\right)^k j(\delta,z)^{-2k}(dz)^k. \end{aligned}$$

Equivalently, for all z in a neighbourhood of ∞,

$$f|_{2k}^{[\delta^{-1}]}(z) = j(\delta^{-1},z)^{-2k} f(\delta^{-1}(z)) = F(e^{2\pi i z/h})\left(\frac{2\pi i}{h}e^{2\pi i z/h}\right)^k,$$

which is to say that f is meromorphic at s. All in all, f is a meromorphic modular form for Γ of weight $2k$.

Conversely, suppose that f is a meromorphic modular form for Γ of weight $2k$. We have to give expressions in local coordinates of a meromorphic k–fold differential form ω on $X(\Gamma)$ which pulls back to $f(z)(dz)^k$. Let $(\Gamma \setminus U, \varphi)$ be a chart

of $X(\Gamma)$ at a point p as defined in section 2.3 and let $v \in U$ such that $\pi(v) = p$.

- If v is neither an elliptic point nor a cusp, ω can be written as $f(z)(dz)^k$ for z in a neighbourhood of v. In particular, ω is meromorphic at p.
- If v is an elliptic point, consider $\delta = \left(\begin{smallmatrix} 1 & -v \\ 1 & -\overline{v} \end{smallmatrix}\right)$ and $d = |\overline{\Gamma_v}|$. Thus, the local coordinate at $\pi(z)$ is $\tau = \delta(z)^d$. Write $w = \delta(z)$. We can express

$$\begin{aligned} f(z)(dz)^k &= f(\delta^{-1}(w))(d\delta^{-1}(w))^k \\ &= \det(\delta^{-1})^k j(\delta^{-1}, w)^{-2k} (f \circ \delta^{-1})(w)(dw)^k = f|_{2k}^{[\delta^{-1}]}(w)(dw)^k, \end{aligned}$$

  and we have that $f|_{2k}^{[\delta^{-1}]}(w)(dw)^k$ is a meromorphic k–fold differential form defined in the unit disc which is invariant under $\delta\Gamma_v\delta^{-1}$ (because $f(z)(dz)^k$ is invariant under Γ_v). Since the elements of $\delta\Gamma_v\delta^{-1}$ correspond to the maps $w \mapsto \zeta_d w$ (where ζ_d is a d–th root of unity), we deduce that

$$f|_{2k}^{[\delta^{-1}]}(w)(dw)^k = f|_{2k}^{[\delta^{-1}]}(\zeta_d w)(d(\zeta_d w))^k = \zeta_d^k f|_{2k}^{[\delta^{-1}]}(\zeta_d w)(dw)^k.$$

  Therefore, $f|_{2k}^{[\delta^{-1}]}(w) = \zeta_d^k f|_{2k}^{[\delta^{-1}]}(\zeta_d w)$ for every d–th root of unity ζ_d, which means that $w^k f|_{2k}^{[\delta^{-1}]}(w)$ has a Laurent series expansion containing only powers of w^d. Now we can write

$$f(z)(dz)^k = f|_{2k}^{[\delta^{-1}]}(\tau^{1/d})(d\tau^{1/d})^k = d^{-k}\tau^{-k}(\tau^{1/d})^k f|_{2k}^{[\delta^{-1}]}(\tau^{1/d})(d\tau)^k,$$

  and this last expression is well-defined and meromorphic at 0 by the previous discussion: we take this to be the local expression of ω at p with respect to φ.
- If v is a cusp, then $(\varphi \circ \pi)(z) = e^{2\pi i \delta(z)/h}$ for all $z \in U$, with $\delta \in \mathrm{SL}_2(\mathbb{Z})$ such that $\delta(v) = \infty$ and $h = [\mathrm{SL}_2(\mathbb{Z})_\infty : \{\pm 1\} \cdot \delta\Gamma_v\delta^{-1}]$. Since f is meromorphic at v, we can write $f|_{2k}^{[\delta^{-1}]}(z) = \widehat{f_v}(e^{2\pi i z/h})$, where $\widehat{f}$ is meromorphic at 0. Now, if $q = e^{2\pi i \delta(z)/h}$, we obtain that

$$\begin{aligned} f(z)(dz)^k &= (f \circ \delta^{-1})(\delta(z)) j(\delta, z)^{2k} \left(\frac{h}{2\pi i q}\, dq\right)^k \\ &= \left(\frac{h}{2\pi i q}\right)^k f|_{2k}^{[\delta^{-1}]}(\delta(z))(dq)^k = \left(\frac{h}{2\pi i q}\right)^k \widehat{f_v}(q)(dq)^k \end{aligned}$$

  because $dq = \frac{2\pi i}{h} j(\delta, z)^{-2} e^{2\pi i \delta(z)/h}\, dz$. Therefore, we take $\left(\frac{h}{2\pi i}\right)^k q^{-k} \widehat{f_v}(q)(dq)^k$ (which is meromorphic) as the local expression of ω at p with respect to φ.

These local expressions are compatible because they are defined so that they pull back to $f(z)(dz)^k$; therefore, ω is well-defined.

One checks easily that the exhibited correspondence preserves linear transformations. □

Definition 2.28. Let k be a positive integer and let ω be a meromorphic k–fold differential form on a Riemann surface M, not identically zero. Let $p \in \mathrm{M}$ and express ω as $f(z)(dz)^k$ in a neighbourhood of p, where z is a local coordinate such that p corresponds to z_0. The *order* of ω at p is the order of f at z_0:

$$\mathrm{ord}_p(\omega) = \mathrm{ord}_{z_0}(f).$$

(This definition is independent of the choice of the local coordinate.)

Lemma 2.29. *Let k be a positive integer. Let f be a meromorphic modular form for Γ of weight $2k$ and let ω be the corresponding meromorphic k–fold differential form on $\mathrm{X}(\Gamma)$. Let $z \in \mathbb{H}^*$ and let $p = \pi(z)$.*

(1) *If z is neither an elliptic point nor a cusp, $\mathrm{ord}_z(f) = \mathrm{ord}_p(\omega)$.*
(2) *If z is an elliptic point, $\mathrm{ord}_z(f) = d_z \,\mathrm{ord}_p(\omega) + k(d_z - 1)$, where $d_z = |\overline{\Gamma_z}|$.*
(3) *If z is a cusp, $\mathrm{ord}_z(f) = \mathrm{ord}_p(\omega) + k$.*

Proof. Let $(\Gamma\backslash \mathrm{U}, \varphi)$ be the chart of $\mathrm{X}(\Gamma)$ at p defined in section 2.3 and consider $\gamma \in \Gamma$ such that $\gamma(z) \in \mathrm{U}$. We have computed explicitly the expression of ω in this local coordinate in the proof of proposition 2.27.

(1) If z is neither an elliptic point nor a cusp, the local coordinate at p is just $\tau = \varphi(p) = \gamma(z)$ and ω can be expressed as $f(\tau)(d\tau)^k$. Since $\gamma \in \Gamma$ and f is weakly modular for Γ, we deduce that $\mathrm{ord}_p(\omega) = \mathrm{ord}_{\gamma(z)}(f) = \mathrm{ord}_z(f)$.
(2) If z is an elliptic point, the local coordinate at p is $\tau = \varphi(p) = \delta(\gamma(z))^{d_z}$ and ω can be written as $d_z^{-k}\tau^{k(-1+1/d_z)} f|_{2k}^{[\delta^{-1}]}(\tau^{1/d_z})(d\tau)^k$. Since $\delta = \begin{pmatrix} 1 & -\gamma(z) \\ 1 & -\overline{\gamma(z)} \end{pmatrix}$ maps $\gamma(z)$ to 0 and $j(\delta^{-1}, 0) = (2i\,\mathrm{Im}(\gamma(z)))^{-1} \notin \{0, \infty\}$, we conclude that

$$\begin{aligned}\mathrm{ord}_p(\omega) &= \frac{\mathrm{ord}_0\left(f|_{2k}^{[\delta^{-1}]}\right)}{d_z} + k\left(\frac{1}{d_z} - 1\right) = \frac{\mathrm{ord}_{\gamma(z)}(f)}{d_z} + k\left(\frac{1}{d_z} - 1\right) \\ &= \frac{\mathrm{ord}_z(f)}{d_z} + k\left(\frac{1}{d_z} - 1\right) = \frac{1}{d_z}[\mathrm{ord}_z(f) - k(d_z - 1)].\end{aligned}$$

(3) If z is a cusp, the local coordinate at p is $q = \varphi(p) = e^{2\pi i \delta(\gamma(z))/h}$ and ω can be expressed as $\left(\frac{h}{2\pi i}\right)^k q^{-k}\widehat{f_{\gamma(z)}}(q)(dq)^k$. Again, since $\gamma \in \Gamma$ and f is weakly modular for Γ, we conclude that $\mathrm{ord}_p(\omega) = \mathrm{ord}_{\gamma(z)}(f) - k = \mathrm{ord}_z(f) - k$. □

Lemma 2.30. *For every positive integer k, there exists a meromorphic modular form for Γ of weight $2k$. Moreover, if g_0 is a meromorphic modular form for Γ of weight $2k$ which is not identically zero, all the others are of the form fg_0, where f is some modular function for Γ.*

Proof. Since $X(\Gamma)$ is a compact Riemann surface, there exists a non-constant meromorphic function F_0 defined on $X(\Gamma)$ (this is a fundamental and non-trivial theorem from the theory of compact Riemann surfaces), and this corresponds to a modular function for Γ, say h_0. In fact, in our case we can take $f_0 = j$ (the modular invariant; see definition 1.23).

Now let $f_1 = f_0'$, which is clearly meromorphic in $\mathbb{H}$. By the chain rule, f_1 is also weakly modular for Γ of weight 2. Finally, if s is a cusp, let $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ such that $\alpha(\infty) = s$ and consider the q_h–expansion $\widehat{f_s}$ of f_0 at s. That is, $f_0(\alpha(z)) = \widehat{f_s}(q_h)$, where $q_h = e^{2\pi i z/h}$. Then,

$$f_1|_2^{[\alpha]}(z) = f_0'(\alpha(z))\alpha'(z) = \widehat{f_s'}(q_h)\frac{2\pi i}{h}q_h,$$

which means that f_1 is also meromorphic at s. All in all, f_1 is a meromorphic modular form for Γ of weight 2. Therefore, one checks easily that $f_k = (f_1)^k$ is a meromorphic modular form for Γ of weight $2k$ (this is analogous to proposition 1.26).

Let g_0 be a meromorphic modular form for Γ of weight $2k$, not identically zero. If f is a modular function for Γ, it is clear that fg_0 is also a meromorphic modular form for Γ of weight $2k$. Conversely, let g be another meromorphic modular form for Γ of weight $2k$ and define $f = \frac{g}{g_0}$. Again, f is meromorphic both in $\mathbb{H}$ and at the cusps (because it is a quotient of meromorphic functions) and $f \circ \gamma = f$ for all $\gamma \in \Gamma$ (because g and g_0 transform in the same way under Γ). Therefore, f is a modular function for Γ. □

Theorem 2.31. *Let g be the genus of $X(\Gamma)$. Let v_2 be the number of inequivalent elliptic points for Γ which are $\mathrm{SL}_2(\mathbb{Z})$–equivalent to i, let v_3 be the number of inequivalent elliptic points for Γ which are $\mathrm{SL}_2(\mathbb{Z})$–equivalent to $\rho = e^{\pi i/3}$ and let v_∞ be the number of inequivalent cusps for Γ. The dimension of $M_{2k}(\Gamma)$ is*

$$\dim(M_{2k}(\Gamma)) = \begin{cases} 0 & \text{if } k < 0, \\ 1 & \text{if } k = 0, \\ (2k-1)(g-1) + \left\lfloor \frac{k}{2} \right\rfloor v_2 + \left\lfloor \frac{2k}{3} \right\rfloor v_3 + k v_\infty & \text{if } k > 0; \end{cases}$$

and the dimension of $S_{2k}(\Gamma)$ *is*

$$\dim(S_{2k}(\Gamma)) = \begin{cases} 0 & \text{if } k \leq 0, \\ g & \text{if } k = 1, \\ (2k-1)(g-1) + \left\lfloor \frac{k}{2} \right\rfloor v_2 + \left\lfloor \frac{2k}{3} \right\rfloor v_3 + (k-1)v_\infty & \text{if } k \geq 2. \end{cases}$$

Proof. Since $X(\Gamma)$ is a compact Riemann surface, a non-constant meromorphic function on $X(\Gamma)$ takes each value in $\mathbb{P}^1_{\mathbb{C}}$ the same number of times (counting multiplicities). But every $f \in M_0(\Gamma)$ corresponds to a holomorphic function on $X(\Gamma)$ (by proposition 2.26) which does not take the value ∞ and so f must be constant. That is, $M_0(\Gamma) = \mathbb{C}$ and $S_0(\Gamma) = \{0\}$.

Suppose that $f \in M_{2k}(\Gamma)$ for some $k < 0$. Then, $f^{12}\Delta^{2k} \in S_0(\Gamma) = \{0\}$, which means that $f = 0$. (Δ is the modular discriminant; see definition 1.20.) Hence, $M_{2k}(\Gamma) = S_{2k}(\Gamma) = \{0\}$ whenever $k < 0$.

In the remainder of the proof, assume that $k > 0$. We compute $\dim(M_{2k}(\Gamma))$ and $\dim(S_{2k}(\Gamma))$ by applying the Riemann–Roch theorem to $X(\Gamma)$. The divisor of a meromorphic function $F\colon X(\Gamma) \to \mathbb{P}^1_{\mathbb{C}}$ is

$$\operatorname{div}(F) = \sum_{p \in X(\Gamma)} \operatorname{ord}_p(F)\, p;$$

similarly, the divisor of a meromorphic k–fold differential form ω on $X(\Gamma)$ is

$$\operatorname{div}(\omega) = \sum_{p \in X(\Gamma)} \operatorname{ord}_p(\omega)\, p.$$

For every divisor $D = \sum_{p \in X(\Gamma)} n_p\, p$ (where $n_p \in \mathbb{Z}$ for all $p \in X(\Gamma)$ and $n_p = 0$ for all but finitely many $p \in X(\Gamma)$), let

$$L(D) = \{0\} \cup \{F\colon X(\Gamma) \to \mathbb{P}^1_{\mathbb{C}} \text{ non-zero and meromorphic} : \operatorname{div}(F) + D \geq 0\}.$$

Define $l(D)$ to be the dimension of $L(D)$ (as a complex vector space) and write $\deg(D) = \sum_{p \in X(\Gamma)} n_p$ (this number is called the degree of D).

It is well-known that a canonical divisor K (i.e., the divisor of a non-zero meromorphic differential form) satisfies that $l(K) = g$ and $\deg(K) = 2g - 2$ (it is a consequence of the Riemann–Roch theorem). Therefore, the degree of a non-zero meromorphic k–fold differential form is $k(2g-2)$ (because it corresponds to the product of k meromorphic differential forms, by proposition 2.27). We fix a

(non-zero) meromorphic k–fold differential form ω_0 (lemma 2.30 guarantees its existence).

Let $f \in \mathrm{M}_{2k}(\Gamma)$ and let ω be the corresponding k–fold differential form on $\mathrm{X}(\Gamma)$. Write $\omega = h\omega_0$, where h is a meromorphic function on $\mathrm{X}(\Gamma)$. By lemma 2.29, for all $z \in \mathbb{H}^*$ with $\pi(z) = p$,

$$0 \leq \operatorname{ord}_z(f) = \begin{cases} d_z \operatorname{ord}_p(\omega) + k(d_z - 1) & \text{if } z \text{ is an elliptic point,} \\ \operatorname{ord}_p(\omega) + k & \text{if } z \text{ is a cusp,} \\ \operatorname{ord}_p(\omega) & \text{otherwise,} \end{cases}$$

or, equivalently (using that $\operatorname{ord}_p(\omega) = \operatorname{ord}_p(h) + \operatorname{ord}_p(\omega_0)$),

$$0 \leq \begin{cases} \operatorname{ord}_p(h) + \operatorname{ord}_p(\omega_0) + k\left(1 - \frac{1}{d_z}\right) & \text{if } z \text{ is an elliptic point,} \\ \operatorname{ord}_p(h) + \operatorname{ord}_p(\omega_0) + k & \text{if } z \text{ is a cusp,} \\ \operatorname{ord}_p(h) + \operatorname{ord}_p(\omega_0) & \text{otherwise.} \end{cases}$$

All the terms appearing in these inequalities are integers except for possibly $k\left(1 - \frac{1}{d_z}\right)$ and so we can replace it with its floor. Adding up the inequalities corresponding to a set of representatives of $\Gamma \backslash \mathbb{H}^*$, we obtain that $\operatorname{div}(h) + \mathrm{D} \geq 0$, where $\mathrm{D} = \operatorname{div}(\omega_0) + \sum_p kp + \sum_p \lfloor k(1 - 1/d_p) \rfloor p$ (here, the first sum is over the images of the cusps and the second sum is over the images of the elliptic points; moreover, $d_p = d_z = |\overline{\Gamma_z}|$ for $z \in \pi^{-1}(p)$). Conversely, if $h \in \mathrm{L}(\mathrm{D})$, the meromorphic modular form corresponding to $h\omega_0$ is, in fact, holomorphic. Therefore, $\mathrm{M}_{2k}(\Gamma)$ and $\mathrm{L}(\mathrm{D})$ are isomorphic (as complex vector spaces). Since $\deg(\mathrm{D}) > 2g - 2$, we have that $\mathrm{L}(\mathrm{K} - \mathrm{D}) = \{0\}$. Thus, by the Riemann–Roch theorem,

$$\begin{aligned} \dim(\mathrm{M}_{2k}(\Gamma)) = l(\mathrm{D}) &= \deg(\mathrm{D}) + 1 - g + l(\mathrm{K} - \mathrm{D}) \\ &= (2k-1)(g-1) + kv_\infty + \left\lfloor \frac{k}{2} \right\rfloor v_2 + \left\lfloor \frac{k}{3} \right\rfloor v_3 . \end{aligned}$$

The dimension of $\mathrm{S}_{2k}(\Gamma)$ can be computed in a similar way. In this case, we have that $\mathrm{S}_{2k}(\Gamma)$ is isomorphic to $\mathrm{L}(\widetilde{\mathrm{D}})$, where $\widetilde{\mathrm{D}} = \mathrm{D} - \sum_p p$ (here, the sum is over the images of the cusps). Indeed, the only inequalities that have changed are those corresponding to cusps, which have become $\operatorname{ord}_p(h) + \operatorname{ord}_p(\omega_0) + k - 1 \geq 0$. If $k \geq 2$, then $\deg(\widetilde{\mathrm{D}}) > 2g - 2$ and $\mathrm{L}(\mathrm{K} - \widetilde{\mathrm{D}}) = \{0\}$, so we can compute $l(\widetilde{\mathrm{D}})$ directly using the Riemann–Roch theorem (as before). Finally, for $k = 1$, $\widetilde{\mathrm{D}} = \operatorname{div}(\omega_0)$ and this is a canonical divisor; thus, $\dim(\mathrm{S}_2(\Gamma)) = l(\widetilde{\mathrm{D}}) = l(\mathrm{K}) = g$. □

There are similar formulae for the dimensions of $M_k(\Gamma)$ and $S_k(\Gamma)$ when k is odd. However, the proofs are a bit more involved and, therefore, we omit them.

Proposition 2.32. *Let g be the genus of $X(\Gamma)$ and let v_∞ be the number of inequivalent cusps for Γ. If f is a meromorphic modular form for Γ of weight $2k$, not identically zero, then*

$$\sum_z \left[\frac{\mathrm{ord}_z(f)}{d_z} - k\left(1 - \frac{1}{d_z}\right)\right] = k(2g-2) + kv_\infty,$$

where the sum is over a set of representatives of $\Gamma \backslash \mathbb{H}^*$ and either $d_z = |\overline{\Gamma_z}|$ if $z \in \mathbb{H}$ or $d_z = 1$ if z is a cusp.

Proof. Let ω be the meromorphic k–fold differential form on $X(\Gamma)$ associated with f. By lemma 2.29, we know that

$$\frac{\mathrm{ord}_z(f)}{d_z} - k\left(1 - \frac{1}{d_z}\right) = \begin{cases} \mathrm{ord}_{\pi(z)}(\omega) & \text{if } z \in \mathbb{H}, \\ \mathrm{ord}_{\pi(z)}(\omega) + k & \text{if } z \text{ is a cusp}. \end{cases}$$

Therefore, summing over a set of representatives of $\Gamma \backslash \mathbb{H}^*$, we obtain that

$$\sum_z \left[\frac{\mathrm{ord}_z(f)}{d_z} - k\left(1 - \frac{1}{d_z}\right)\right] = \deg(\mathrm{div}(\omega)) + kv_\infty = k(2g-2) + kv_\infty.$$

□

All these formulae generalise results which were proved in a more elementary way for $SL_2(\mathbb{Z})$ in chapter 1.

Chapter 3

Hecke operators

Hecke operators are averaging operators acting on the space of modular forms. They are multiplicative and satisfy certain recurrence relations. These properties can be summarised in certain formal Euler products. One can then show that the coefficients of the Fourier expansions of Hecke eigenforms (i.e., eigenvectors of the Hecke operators) satisfy similar relations.

This chapter presents the basic theory of Hecke operators for some special congruence subgroups (including those of interest to us). We prove that Hecke operators are self-adjoint with respect to a certain Hermitian inner product on the spaces of cusp forms and, consequently, there are bases of Hecke eigenforms for these spaces.

The contents of this chapter are based principally on the sections dedicated to Hecke operators of Koblitz's book [3] and Milne's notes [8].

3.1 The Petersson inner product

The upper half-plane $\mathbb{H}$ can be regarded as a model for hyperbolic plane geometry. The corresponding metric (called the Poincaré metric) is given by the tensor

$$(ds)^2 = \frac{(dx)^2 + (dy)^2}{y^2} = \frac{dz\, d\overline{z}}{y^2},$$

where we write $z = x + iy$ with $x, y \in \mathbb{R}$. This metric induces a volume form

$$\Omega = y^{-2}\, dx \wedge dy = \frac{i}{2}(\mathrm{Im}(z))^{-2}\, dz \wedge d\overline{z},$$

which is invariant under $\mathrm{GL}_2^+(\mathbb{R})$. Indeed, for all $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{R})$, a straightforward computation yields

$$d\alpha(z) = \frac{\det(\alpha)}{(cz+d)^2}\, dz, \quad d\overline{\alpha(z)} = \frac{\det(\alpha)}{(c\overline{z}+d)^2}\, d\overline{z} \quad \text{and} \quad \mathrm{Im}(\alpha(z)) = \frac{\det(\alpha)}{|cz+d|^2}\, \mathrm{Im}(z);$$

hence, $\frac{i}{2}(\mathfrak{Im}(\alpha(z)))^{-2}\,d\alpha(z)\wedge d\overline{\alpha(z)} = \frac{i}{2}(\mathfrak{Im}(z))^{-2}\,dz\wedge d\bar{z}$. On the other hand, since the set of cusps $\mathbb{P}^1_{\mathbb{Q}}$ is a countable set, it has measure zero and so Ω can be used to integrate over subsets of the extended upper half-plane $\mathbb{H}^*$.

Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Consider the fundamental domain $\mathrm{F} = \left\{z\in\mathbb{H} : |z|\geq 1 \text{ and } |\mathfrak{Re}(z)|\leq \frac{1}{2}\right\}$ for $\mathrm{SL}_2(\mathbb{Z})$. Also, let $\alpha_1,\ldots,\alpha_n \in \mathrm{SL}_2(\mathbb{Z})$ be representatives of the left cosets of $\overline{\Gamma}$ in $\mathrm{PSL}_2(\mathbb{Z})$ so that

$$\mathrm{F}_\Gamma = \bigcup_{j=1}^{n} \alpha_j^{-1}(\mathrm{F})$$

is a fundamental domain for Γ (by proposition 1.10).

Proposition 3.1. *Define*

$$\mu(\Gamma) = \int_{\mathrm{F}_\Gamma} \frac{dx\wedge dy}{y^2}.$$

(1) *The integral which defines* $\mu(\Gamma)$ *converges and is independent of the choice of the fundamental domain* F_Γ.
(2) $n = [\mathrm{PSL}_2(\mathrm{Z}) : \overline{\Gamma}] = \mu(\Gamma)/\mu(\mathrm{SL}_2(\mathbb{Z}))$.
(3) *If* $\alpha\in\mathrm{GL}_2^+(\mathbb{Q})$ *and* $\alpha^{-1}\Gamma\alpha\subseteq\mathrm{SL}_2(\mathbb{Z})$, *then* $[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}] = [\mathrm{PSL}_2(\mathbb{Z}) : \alpha^{-1}\overline{\Gamma}\alpha]$.

Proof. First, observe that

$$\mu(\mathrm{SL}_2(\mathbb{Z})) = \int_{\mathrm{F}} y^{-2}\,dx\wedge dy = \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{\sqrt{1-x^2}}^{\infty} y^{-2}\,dy\,dx = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{dx}{\sqrt{1-x^2}} = \frac{\pi}{3} < \infty$$

(where the last integral can be evaluated using the change of variables $x = \sin(t)$). Now, for each $j\in\{1,\ldots,n\}$, we make the change of variables $z\mapsto\alpha_j(z)$:

$$\int_{\alpha_j^{-1}(\mathrm{F})} y^{-2}\,dx\wedge dy = \int_{\mathrm{F}} y^{-2}\,dx\wedge dy.$$

Therefore, the integral which defines $\mu(\Gamma)$ converges and, if μ is well-defined, we have that $\mu(\Gamma) = n\mu(\mathrm{SL}_2(\mathbb{Z}))$.

Moreover, if F'_Γ is another fundamental domain for Γ, we can divide it into regions R_i such that $\beta_i(\mathrm{R}_i)\subseteq\mathrm{F}_\Gamma$ for some $\beta_i\in\Gamma$. In this situation, the invariance of $y^{-2}\,dx\wedge dy$ implies that the integrals over R_i and over $\beta_i(\mathrm{R}_i)$ coincide. This proves that $\mu(\Gamma)$ is well-defined.

Finally, observe that $\alpha^{-1}(\mathrm{F}_\Gamma)$ is a fundamental domain for $\alpha^{-1}\Gamma\alpha$. Therefore, $\mu(\alpha^{-1}\Gamma\alpha) = \mu(\Gamma)$. □

Let k be a positive integer and let $f, g \in \mathrm{M}_k(\Gamma)$.

Lemma 3.2. *Let* $\mathrm{H}(f,g)(z) = f(z)\overline{g(z)}(\mathrm{Im}(z))^k$. *For all* $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$,

$$\mathrm{H}(f,g) \circ \alpha = \mathrm{H}\Big(f|_k^{[\alpha]}, g|_k^{[\alpha]} \Big).$$

Proof. A straight-forward computation yields

$$\begin{aligned} \mathrm{H}(f,g)(\alpha(z)) &= f(\alpha(z))\overline{g(\alpha(z))}(\mathrm{Im}(\alpha(z)))^k = f(\alpha(z))\overline{g(\alpha(z))}\left[\frac{\det(\alpha)}{|j(\alpha,z)|^2}\mathrm{Im}(z)\right]^k \\ &= f|_k^{[\alpha]}(z)\overline{g|_k^{[\alpha]}(z)}(\mathrm{Im}(z))^k = \mathrm{H}\Big(f|_k^{[\alpha]}, g|_k^{[\alpha]} \Big)(z), \end{aligned}$$

as claimed. □

Lemma 3.3. *If at least one of* f *and* g *is a cusp form, then the integral*

$$\int_{\mathrm{F}_\Gamma} f(z)\overline{g(z)} y^k \frac{dx \wedge dy}{y^2}$$

converges absolutely and does not depend on the choice of the fundamental domain F_Γ. *(Here, we write* $z = x + iy$ *with* $x, y \in \mathbb{R}$.*)*

Proof. In each region $\alpha_j^{-1}(\mathrm{F})$, we make the change of variables $z \mapsto \alpha_j^{-1}(z)$ and, by lemma 3.2, the integral becomes

$$\sum_{j=1}^{n} \int_{\mathrm{F}} f|_k^{[\alpha_j^{-1}]}(z)\overline{g|_k^{[\alpha_j^{-1}]}(z)} y^k \frac{dx \wedge dy}{y^2}.$$

Since $f|_k^{[\alpha_j^{-1}]}, g|_k^{[\alpha_j^{-1}]} \in \mathrm{M}_k(\alpha_j^{-1}\Gamma\alpha_j)$, we can express

$$f|_k^{[\alpha_j^{-1}]}(z) = \sum_{n=0}^{\infty} a_n q_{h_j}^n \quad \text{and} \quad g|_k^{[\alpha_j^{-1}]}(z) = \sum_{n=0}^{\infty} b_n q_{h_j}^n,$$

where $q_{h_j} = e^{2\pi i z/h_j}$, for some $h_j \in \mathbb{N}$. These q_{h_j}–expansions are holomorphic at 0 and, by hypothesis, $a_0 = 0$ or $b_0 = 0$. Hence,

$$f|_k^{[\alpha_j^{-1}]}(z)\overline{g|_k^{[\alpha_j^{-1}]}(z)} = q_{h_j}\psi_j(q_{h_j})$$

for some function ψ_j holomorphic at 0, which implies that $|q_{h_j}\psi_j(q_{h_j})| \ll e^{-cy}$ for

some constant $c > 0$ (if y is sufficiently large). But $\int_F e^{-cy} y^{k-2}\, dx\, dy < \infty$, so the integral in the statement of the lemma is absolutely convergent.

Now suppose that F'_Γ is another fundamental domain for Γ. We can divide F'_Γ into regions R_i such that $\beta_i(R_i) \subseteq F_\gamma$ with $\beta_i \in \Gamma$ and, since $f(z)\overline{g(z)}y^k y^{-2}\, dx \wedge dy$ is invariant under Γ, the integrals over R_i and over $\beta_i(R_i)$ coincide. That is, the integral in the statement of the lemma is independent of the choice of F_Γ. □

From here on, assume further that at least one of f and g is a cusp form.

Definition 3.4. The *Petersson inner product* of f and g is defined to be

$$\langle f, g\rangle_\Gamma = \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}]} \int_{F_\Gamma} f(z)\overline{g(z)}y^k \frac{dx \wedge dy}{y^2},$$

where $z = x + iy$. It is immediate from this definition that

$$\begin{aligned} \langle \cdot, \cdot \rangle_\Gamma \colon S_k(\Gamma) \times S_k(\Gamma) &\longrightarrow \mathbb{C} \\ (f_1, f_2) &\longmapsto \langle f_1, f_2\rangle_\Gamma \end{aligned}$$

is a Hermitian inner product on $S_k(\Gamma)$; that is to say, $\langle \cdot, \cdot \rangle_\Gamma$ is:

(i) linear in the first variable and antilinear in the second;
(ii) antisymmetric;
(iii) positive definite.

Since $y^{-2}\, dx \wedge dy$ is invariant under Γ, we could use it to define a measure on $X(\Gamma)$ (see section 2.5 of Shimura's book [12] for the details). Lemma 3.2 shows that $f(z)\overline{g(z)}(\mathrm{Im}(z))^k$ induces a function on $X(\Gamma)$; thus, $\langle f, g\rangle_\Gamma$ corresponds (up to the constant factor $[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}]$) to the integral of this function over $X(\Gamma)$. That is why we may write

$$\langle f, g\rangle_\Gamma = \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}]} \int_{X(\Gamma)} f(z)\overline{g(z)}y^k \frac{dx \wedge dy}{y^2}$$

when we do not want to make the choice of the fundamental domain explicit. This notation is justified by lemma 3.3.

In the following sections, we are going to define operators $T(n)$ for all $n \in \mathbb{N}$ acting on $M_k(\Gamma)$. We are going to prove that (most of) these operators are self-adjoint with respect to the Petersson inner product. Our interest will then lie in the computation of a basis of $S_k(\Gamma)$ consisting of cusp forms which are eigenvectors (eigenforms) of all these operators.

Lemma 3.5. *Let* V *be a finite-dimensional vector space over* $\mathbb{C}$ *with a Hermitian inner product* $\langle \cdot, \cdot \rangle$.

(1) *If* $\varphi\colon V \to V$ *is a self-adjoint linear map, then* V *has a basis consisting of mutually orthogonal eigenvectors of* φ.

(2) *Let* $\varphi_1, \varphi_2, \ldots$ *be a sequence of commuting self-adjoint endomorphisms. There is a basis of* V *consisting of vectors which are eigenvectors of every* φ_n *(*$n \in \mathbb{N}$*). (That is,* $\varphi_1, \varphi_2, \ldots$ *are simultaneously diagonalisable.)*

Proof. If $V = \{0\}$, the result is trivial. Hence, assume that $\dim(V) = d \geq 1$.

Every endomorphism of V has at least one eigenvalue because $\mathbb{C}$ is algebraically closed (and so the characteristic polynomial has at least one root). Therefore, φ has an eigenvector e_1. If $d = 1$, we are done. Otherwise, let $V_1 = (\mathbb{C}e_1)^\perp$. Since φ is self-adjoint, V_1 is stable under φ and we can apply the same argument to $\varphi\big|_{V_1}$ to obtain another eigenvector e_2. Now let $V_2 = (\mathbb{C}e_1 + \mathbb{C}e_2)^\perp$ and we can continue in this manner until we obtain a basis $e_1, e_2, \ldots, e_d$ of V.

If $\varphi_1, \varphi_2, \ldots$ are commuting self-adjoint endomorphisms, we can express $V = \bigoplus_i V(\lambda_i)$, where the λ_i are the distinct eigenvalues of φ_1 and the $V(\lambda_i)$ are the corresponding eigenspaces. For all $n \in \mathbb{N}$, since φ_n commutes with φ_1, φ_n preserves each $V(\lambda_i)$: if $v \in V(\lambda_i)$, then $\varphi_1(\varphi_n(v)) = \varphi_n(\varphi_1(v)) = \lambda_i \varphi_n(v)$ and so $\varphi_n(v) \in V(\lambda_i)$ as well. Therefore, we can decompose each $V(\lambda_i)$ further into a sum of eigenspaces of φ_2; then, we decompose the resulting spaces into a sum of eigenspaces of φ_3, and so on. Since V is finite-dimensional, after finitely many steps this process must stop giving us any new smaller spaces. The resulting decomposition is $V = \bigoplus_i V_i$ such that each φ_n acts as a scalar on each V_i. Now choose a basis for each V_i and take their union. □

Finally, we study some properties of $\langle \cdot, \cdot \rangle_\Gamma$ which are going to be useful after we define the Hecke operators.

Proposition 3.6. *If* Γ' *is another congruence subgroup of* $\mathrm{SL}_2(\mathbb{Z})$ *and* $f, g \in M_k(\Gamma')$ *as well, then* $\langle f, g \rangle_\Gamma = \langle f, g \rangle_{\Gamma'}$.

Proof. First assume that $\Gamma' \subseteq \Gamma$. In this case, consider a set of representatives $\beta_1, \ldots, \beta_m$ of the left cosets of $\overline{\Gamma'}$ in $\overline{\Gamma}$ and choose

$$F_{\Gamma'} = \bigcup_{j=1}^{m} \beta_j^{-1}(F_\Gamma)$$

as a fundamental domain for Γ'. Since $f, g \in \mathrm{M}_k(\Gamma)$, we have that $f|_k^{[\beta_j^{-1}]} = f$ and $g|_k^{[\beta_j^{-1}]} = g$ for all j. Therefore,

$$\begin{aligned}\langle f, g\rangle_{\Gamma'} &= \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma'}]} \sum_{j=1}^{m} \int_{\beta_j^{-1}(\mathrm{F}_\Gamma)} f(z)\overline{g(z)} y^k \frac{dx \wedge dy}{y^2} \\ &= \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}] \cdot [\overline{\Gamma} : \overline{\Gamma'}]} \sum_{j=1}^{m} \int_{\mathrm{F}_\Gamma} f(z)\overline{g(z)} y^k \frac{dx \wedge dy}{y^2} = \langle f, g\rangle_{\Gamma}.\end{aligned}$$

In general, we define $\Gamma'' = \Gamma \cap \Gamma'$ and, by the case proved in the previous paragraph, $\langle f, g\rangle_\Gamma = \langle f, g\rangle_{\Gamma''} = \langle f, g\rangle_{\Gamma'}$. □

Since $\langle f, g\rangle_\Gamma$ is independent of the choice of Γ, we can omit the subgroup from the notation and write $\langle f, g\rangle = \langle f, g\rangle_\Gamma$ as long as there is a sufficiently small congruence subgroup Γ so that $f, g \in \mathrm{M}_k(\Gamma)$.

Lemma 3.7. *Let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ and set $\Gamma' = \alpha^{-1}\Gamma\alpha \cap \mathrm{SL}_2(\mathbb{Z})$. Then, Γ' is a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and the map $f \mapsto f|_k^{[\alpha]}$ takes $\mathrm{M}_k(\Gamma)$ to $\mathrm{M}_k(\Gamma')$ and $\mathrm{S}_k(\Gamma)$ to $\mathrm{S}_k(\Gamma')$.*

Proof. Observe that α can be multiplied by a positive scalar without affecting the action of $|_k^{[\alpha]}$. Therefore, we may assume that α has integer entries.

Let $\mathrm{D} = \det(\alpha)$ and suppose that $\Gamma(\mathrm{N}) \subseteq \Gamma$ for some $\mathrm{N} \in \mathbb{N}$. We are going to prove that $\Gamma(\mathrm{ND}) \subseteq \Gamma'$. Indeed, for all $\gamma \in \Gamma(\mathrm{ND})$, we can write $\gamma = 1 + \mathrm{ND}\beta$ for some matrix β with integer coefficients. Hence, $\alpha\gamma\alpha^{-1} = 1 + \mathrm{N}\alpha\beta(\mathrm{D}\alpha^{-1}) \in \Gamma(\mathrm{N})$ (note that $\det(\alpha\gamma\alpha^{-1}) = \det(\alpha) = 1$), which implies that $\gamma \in \alpha^{-1}\Gamma(\mathrm{N})\alpha \subseteq \alpha^{-1}\Gamma\alpha$ and so $\gamma \in \Gamma'$ too.

It is clear that, if $f \in \mathrm{M}_k(\Gamma)$, then $f|_k^{[\alpha]}$ is weakly modular for Γ' of weight k and holomorphic in $\mathbb{H}$. For all $\beta \in \mathrm{SL}_2(\mathbb{Z})$, it is possible to put $\alpha\beta$ into upper triangular form by using elementary operations of the following types: adding a multiple of one row to another and swapping two rows. Therefore, there exists $\gamma_0 \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma_0^{-1}\alpha\beta = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$, where a and d are positive. If

$$f|_k^{[\gamma_0]}(z) = \sum_{n=n_0}^{\infty} x_n e^{2\pi i n z/h},$$

then we have a Fourier expansion of $f|_k^{[\alpha]}$ at $\beta(\infty)$ given by

$$\left(f|_k^{[\alpha]}\right)\Big|_k^{[\beta]}(z) = (ad)^{\frac{k}{2}} d^{-k} \sum_{n=n_0}^{\infty} x_n e^{2\pi i n(az+b)/(dh)} = \sum_{n=an_0}^{\infty} y_n e^{2\pi i n z/(dh)},$$

where

$$y_n = \begin{cases} 0 & \text{if } a \nmid n, \\ \left(\frac{a}{d}\right)^{\frac{k}{2}} x_{n/a}\, e^{2\pi i bn/(adh)} & \text{if } a \mid n. \end{cases}$$

Thus, if $f \in \mathrm{M}_k(\Gamma)$ (resp. $f \in \mathrm{S}_k(\Gamma)$), then $f|_k^{[\alpha]} \in \mathrm{M}_k(\Gamma')$ (resp. $f|_k^{[\alpha]} \in \mathrm{S}_k(\Gamma')$). □

Proposition 3.8. *Let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ and let $\alpha' = \det(\alpha)\alpha^{-1}$.*

(1) $\langle f|_k^{[\alpha]}, g|_k^{[\alpha]}\rangle = \langle f, g\rangle$.
(2) $\langle f|_k^{[\alpha]}, g\rangle = \langle f, g|_k^{[\alpha']}\rangle$.
(3) $\langle f|_k^{[\alpha]}, g\rangle$ *depends only on the double coset* $\Gamma\alpha\Gamma$. *That is to say, for all* $\gamma_1, \gamma_2 \in \Gamma$, $\langle f|_k^{[\gamma_1\alpha\gamma_2]}, g\rangle = \langle f|_k^{[\alpha]}, g\rangle$.

Proof. Assume (up to multiplication by a suitable integer) that α has integer entries (and so α' too).

Let $\Gamma' = \Gamma \cap \alpha\Gamma\alpha^{-1}$. Γ' is a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ because it is the intersection of the congruence subgroups Γ and $\mathrm{SL}_2(\mathbb{Z}) \cap \alpha\Gamma\alpha^{-1}$. We also know that $f, g \in \mathrm{M}_k(\Gamma) \subseteq \mathrm{M}_k(\Gamma')$ and $f|_k^{[\alpha]}, g|_k^{[\alpha]} \in \mathrm{M}_k(\alpha^{-1}\Gamma\alpha \cap \mathrm{SL}_2(\mathbb{Z})) \subseteq \mathrm{M}_k(\alpha^{-1}\Gamma'\alpha)$, by lemma 3.7. On the other hand, if F' is a fundamental domain for Γ', $\alpha^{-1}(\mathrm{F}')$ is a fundamental domain for $\alpha^{-1}\Gamma'\alpha$. Therefore,

$$\begin{aligned} \langle f|_k^{[\alpha]}, g|_k^{[\alpha]}\rangle &= \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \alpha^{-1}\overline{\Gamma'}\alpha]} \int_{\alpha^{-1}(\mathrm{F}')} f|_k^{[\alpha]}(z)\overline{g|_k^{[\alpha]}(z)}\, y^k \frac{dx \wedge dy}{y^2} \\ &= \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma'}]} \int_{\mathrm{F}'} f(z)\overline{g(z)} y^k \frac{dx \wedge dy}{y^2} = \langle f, g\rangle \end{aligned}$$

because $[\mathrm{PSL}_2(\mathbb{Z}) : \alpha^{-1}\overline{\Gamma'}\alpha] = [\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma'}]$ (by proposition 3.1).

Applying the previous result with g replaced by $g|_k^{[\alpha^{-1}]}$ (which is the same as $g|_k^{[\alpha']}$), we obtain that $\langle f|_k^{[\alpha]}, g\rangle = \langle f, g|_k^{[\alpha^{-1}]}\rangle = \langle f, g|_k^{[\alpha']}\rangle$. And, since $f, g \in \mathrm{M}_k(\Gamma)$, $\langle f|_k^{[\gamma_1\alpha\gamma_2]}, g\rangle = \left\langle \left(f|_k^{[\gamma_1]}\right)\Big|_k^{[\alpha]}, g|_k^{[\gamma_2^{-1}]}\right\rangle = \langle f|_k^{[\alpha]}, g\rangle$ for all $\gamma_1, \gamma_2 \in \Gamma$. □

3.2 Hecke operators for $\mathrm{SL}_2(\mathbb{Z})$

Hecke operators play a fundamental role in the theory of modular forms. In this section, we introduce Hecke operators for the full modular group and briefly explain some basic facts.

Hecke operators were first introduced in order to study some properties of the Ramanujan τ–function, which is related to the q–expansion of the modular

discriminant $\Delta(z)$. These results, which are omitted from this work, are explained in detail in Serre's book [11] and in Milne's notes [8], for instance.

Recall that a lattice in $\mathbb{C}$ is a subgroup of the form $\Lambda = \Lambda(\omega_1, \omega_2) = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$ where ω_1 and ω_2 are complex numbers which are linearly independent over $\mathbb{R}$. We shall always assume (up to interchanging ω_1 and ω_2) that $\operatorname{Im}\left(\frac{\omega_1}{\omega_2}\right) > 0$. Let $\mathcal{L}$ denote the set of lattices in $\mathbb{C}$. One checks easily that $\Lambda(\omega_1, \omega_2) = \Lambda(\omega_1', \omega_2')$ if and only if $\omega_1' = a\omega_1 + b\omega_2$ and $\omega_2' = c\omega_1 + d\omega_2$ for some $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$.

Every lattice Λ has associated with it an elliptic curve $\mathrm{E}_\Lambda = \mathbb{C}/\Lambda$. One can show that two elliptic curves $\mathbb{C}/\Lambda$ and $\mathbb{C}/\Lambda'$ are isomorphic if and only if $\Lambda = \lambda\Lambda'$ for some $\lambda \in \mathbb{C}^\times$. Therefore, we consider the (left) action of $\mathbb{C}^\times$ on $\mathcal{L}$ by homotheties and a lattice $\Lambda(\omega_1, \omega_2)$ is equivalent to $\Lambda(\tau, 1)$ in $\mathbb{C}^\times \backslash \mathcal{L}$, where $\tau = \frac{\omega_1}{\omega_2} \in \mathbb{H}$. Thus, we write $\Lambda(\tau) = \Lambda(\tau, 1)$ for all $\tau \in \mathbb{H}$. We see that $\Lambda(\tau)$ and $\Lambda(\tau')$ coincide in $\mathbb{C}^\times \backslash \mathcal{L}$ if and only if there exists $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that $\tau' = \frac{a\tau+b}{c\tau+d}$. Hence, there is a bijective correspondence between the elements of $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}$ and the isomorphism classes of elliptic curves. $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}$ is said to be a moduli space of elliptic curves.

Throughout this section, k will be an integer.

Lemma 3.9. *Let* $\mathrm{F}\colon \mathcal{L} \to \mathbb{P}^1_{\mathbb{C}}$ *be a homogeneous function of degree* $-2k$ *(that is, such that* $\mathrm{F}(\lambda\Lambda) = \lambda^{-2k}\mathrm{F}(\Lambda)$ *for all* $\lambda \in \mathbb{C}^\times$ *and all* $\Lambda \in \mathcal{L}$*). The function*

$$\begin{aligned} f\colon \mathbb{H} &\longrightarrow \mathbb{P}^1_{\mathbb{C}} \\ z &\longmapsto f(z) = \mathrm{F}(\Lambda(z, 1)) \end{aligned}$$

is weakly modular for $\mathrm{SL}_2(\mathbb{Z})$ *of weight* $2k$*. Moreover, the map* $\mathrm{F} \mapsto f$ *is a bijection between functions of lattices which are homogeneous of degree* $-2k$ *and weakly modular functions for* $\mathrm{SL}_2(\mathbb{Z})$ *of weight* $2k$*.*

Proof. We write $\mathrm{F}(\omega_1, \omega_2) = \mathrm{F}(\Lambda(\omega_1, \omega_2))$. Note that $\mathrm{F}(\lambda\omega_1, \lambda\omega_2) = \lambda^{-2k}\mathrm{F}(\omega_1, \omega_2)$ for all $\lambda \in \mathbb{C}^\times$ and $\mathrm{F}(a\omega_1 + b\omega_2, c\omega_1 + d\omega_2) = \mathrm{F}(\omega_1, \omega_2)$ for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. We see thus that the product $\omega_2^{2k}\,\mathrm{F}(\omega_1, \omega_2)$ depends only on $\frac{\omega_1}{\omega_2}$ and, consequently, there is a function $f(z)$ such that $\mathrm{F}(\omega_1, \omega_2) = \omega_2^{-2k} f\left(\frac{\omega_1}{\omega_2}\right)$. Thus, for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$,

$$(c\omega_1 + d\omega_2)^{-2k} f\left(\frac{a\omega_1 + b\omega_2}{c\omega_1 + d\omega_2}\right) = \omega_2^{-2k} f\left(\frac{\omega_1}{\omega_2}\right)$$

or, equivalently, $(cz + d)^{-2k} f\left(\frac{az+b}{cz+d}\right) = f(z)$.

Conversely, given a weakly modular function f for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$, we can define $\mathrm{F}(\omega_1, \omega_2) = \omega_2^{-2k} f\left(\frac{\omega_1}{\omega_2}\right)$, which is homogeneous of degree $-2k$. □

We are going to use this interpretation of modular forms to define Hecke operators. To this aim, we first define operators on $\mathcal{L}$, which define operators on functions on $\mathcal{L}$.

Let $\mathscr{L}$ be the free abelian group generated by the elements of $\mathcal{L}$.

Definition 3.10. For every $n \in \mathbb{N}$, the *Hecke operator* $\mathrm{T}(n)\colon \mathscr{L} \to \mathscr{L}$ is the only $\mathbb{Z}$–linear operator which maps each lattice Λ to the sum of all of its sublattices of index n:

$$\mathrm{T}(n)\Lambda = \sum_{[\Lambda:\Lambda']=n} \Lambda'.$$

(This sum is finite because any such sublattice Λ' contains $n\Lambda$ and $\Lambda/n\Lambda$ is finite.)

We also consider the homothety operators $\mathrm{R}(n)\colon \mathscr{L} \to \mathscr{L}$, which are the linear maps defined by $\mathrm{R}(n)\Lambda = n\Lambda$ for all $\Lambda \in \mathcal{L}$.

Proposition 3.11. *The Hecke operators and the homothety operators (as endomorphisms of $\mathscr{L}$) satisfy the following identities:*

(1) $\mathrm{R}(m) \circ \mathrm{R}(n) = \mathrm{R}(n) \circ \mathrm{R}(m) = \mathrm{R}(mn)$ *for all* $m, n \in \mathbb{N}$;
(2) $\mathrm{R}(m) \circ \mathrm{T}(n) = \mathrm{T}(n) \circ \mathrm{R}(m)$ *for all* $m, n \in \mathbb{N}$;
(3) $\mathrm{T}(m) \circ \mathrm{T}(n) = \mathrm{T}(n) \circ \mathrm{T}(m) = \mathrm{T}(mn)$ *for all* $m, n \in \mathbb{N}$ *such that* $(m, n) = 1$;
(4) $\mathrm{T}(p^n) \circ \mathrm{T}(p) = \mathrm{T}(p^{n+1}) + p\,\mathrm{R}(p) \circ \mathrm{T}(p^{n-1})$ *for all prime* p *and all* $n \in \mathbb{N}$.

Proof. The first two identities are trivial.

To prove the third identity, fix a lattice Λ. For every sublattice Λ'' of Λ of index mn, there exists a unique sublattice Λ' of Λ containing Λ'' and such that $[\Lambda : \Lambda'] = n$ and $[\Lambda' : \Lambda''] = m$. Indeed, Λ/Λ'' is an abelian group of order mn which decomposes uniquely as the direct sum of a group of order m and a group of order n (because $(m, n) = 1$). Therefore, $\mathrm{T}(mn)\Lambda = (\mathrm{T}(m) \circ \mathrm{T}(n))\Lambda$.

Finally, we prove the last identity with a similar argument. Let Λ be a lattice. We observe that $(\mathrm{T}(p^n) \circ \mathrm{T}(p))\Lambda$, $\mathrm{T}(p^{n+1})\Lambda$ and $(p\,\mathrm{R}(p) \circ \mathrm{T}(p^{n-1}))\Lambda$ are all sums of sublattices of Λ of index p^{n+1}. One such sublattice Λ'' occurs exactly a times in the first sum, exactly once in the second sum and exactly b times in the third sum, so we have to prove that $a = 1 + pb$. To do so, we distinguish two cases.

If Λ'' is not contained in $p\Lambda$, it is clear that $b = 0$. On the other hand, a is the number of sublattices Λ' of Λ containing Λ'' and of index p in Λ. Such a lattice

Λ' contains $p\Lambda$, and its image in $\Lambda/p\Lambda$ is of order p and contains the image of Λ'' (which is also of order p). Thus, there is exactly one possible Λ' with these properties, which means that $a = 1$.

If Λ'' is contained in $p\Lambda$, we have that $b = 1$. But every sublattice Λ' of Λ of index p contains $p\Lambda$ and so Λ'' too. Therefore, a coincides with the number of sublattices of Λ of index p (or, equivalently, with the number of subgroups of $\Lambda/p\Lambda \simeq (\mathbb{Z}/p\mathbb{Z})^2$ of index p), and this is $\frac{p^2-1}{p-1} = p+1$. □

Corollary 3.12. *The $\mathbb{Z}$–algebra generated by the $\mathrm{T}(p)$ and $\mathrm{R}(p)$ for p prime is commutative and contains all the $\mathrm{T}(n)$ for $n \in \mathbb{N}$.*

Definition 3.13. There is an action of Hecke operators and homothety operators on the set of functions $\mathrm{F}\colon \mathcal{L} \to \mathbb{P}^1_{\mathbb{C}}$ which are homogeneous of degree $-2k$: for all $n \in \mathbb{N}$, we define $\mathrm{T}(n)\mathrm{F}$ and $\mathrm{R}(n)\mathrm{F}$ to be the functions given by

$$\mathrm{T}(n)\mathrm{F}(\Lambda) = \mathrm{F}(\mathrm{T}(n)\Lambda) = \sum_{[\Lambda:\Lambda']=n} \mathrm{F}(\Lambda') \quad \text{and} \quad \mathrm{R}(n)\mathrm{F}(\Lambda) = \mathrm{F}(\mathrm{R}(n)\Lambda) = n^{-2k}\mathrm{F}(\Lambda)$$

for all $\Lambda \in \mathcal{L}$.

Proposition 3.14. *Let $\mathrm{F}\colon \mathcal{L} \to \mathbb{P}^1_{\mathbb{C}}$ be a homogeneous function of degree $-2k$. For all $n \in \mathbb{N}$, $\mathrm{T}(n)\mathrm{F}\colon \mathcal{L} \to \mathbb{P}^1_{\mathbb{C}}$ is also homogeneous of degree $-2k$. Moreover,*

(1) $\mathrm{T}(m)\mathrm{T}(n)\mathrm{F} = \mathrm{T}(mn)\mathrm{F}$ *for all $m, n \in \mathbb{N}$ such that $(m,n) = 1$, and*

(2) $\mathrm{T}(p)\mathrm{T}(p^n)\mathrm{F} = \mathrm{T}(p^{n+1})\mathrm{F} + p^{1-2k}\mathrm{T}(p^{n-1})\mathrm{F}$ *for all prime p and all $n \in \mathbb{N}$.*

Proof. It is immediate from the definition of $\mathrm{T}(n)\mathrm{F}$ and proposition 3.11. □

Definition 3.15. Let $f\colon \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ be a weakly modular function for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$ and let $\mathrm{F}\colon \mathcal{L} \to \mathbb{P}^1_{\mathbb{C}}$ be the associated homogeneous function of degree $-2k$ (as in lemma 3.9). For every $n \in \mathbb{N}$, we define $\mathrm{T}(n)f\colon \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ to be the function associated with $n^{2k-1}\mathrm{T}(n)\mathrm{F}$:

$$\mathrm{T}(n)f(z) = n^{2k-1}\mathrm{T}(n)\mathrm{F}(\Lambda(z,1)).$$

(The factor n^{2k-1} is introduced so that some formulae have integer coefficients.)

Proposition 3.16. *If $f\colon \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ is weakly modular for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$, then $\mathrm{T}(n)f\colon \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ is also weakly modular for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$ for every $n \in \mathbb{N}$. Moreover,*

(1) $\mathrm{T}(m)\mathrm{T}(n)f = \mathrm{T}(mn)f$ *for all $m, n \in \mathbb{N}$ such that $(m,n) = 1$, and*

(2) $\mathrm{T}(p)\,\mathrm{T}(p^n)\,f = \mathrm{T}(p^{n+1})\,f + p^{2k-1}\,\mathrm{T}(p^{n-1})\,f$ *for all prime* p *and all* $n \in \mathbb{N}$.

Proof. It is immediate from the definition of $\mathrm{T}(n)\,f$ and proposition 3.14 (taking into account the additional factor). □

We have defined the action of Hecke operators on weakly modular functions and so, in particular, on modular forms. Nevertheless, the definition is quite abstract and the properties of the functions obtained in this way are not obvious. The following results provide simpler descriptions of these functions and even precise formulae to compute them.

Lemma 3.17. *Let* A *be a* 2×2 *matrix with entries in* $\mathbb{Z}$ *and* $\det(\mathrm{A}) = n \neq 0$. *There exists* $\mathrm{U} \in \mathrm{SL}_2(\mathbb{Z})$ *such that* $\mathrm{UA} = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ *with* $ad = n$, $a \geq 1$ *and* $0 \leq b < d$. *Moreover, the integers* a, b *and* d *are uniquely determined.*

Proof. It is possible to put A into upper triangular form by using elementary operations of the following types: adding a multiple of one row to another and swapping two rows. Since these operations are invertible, they correspond to left multiplication by a matrix in $\mathrm{SL}_2(\mathbb{Z})$. We can assume, up to multiplication by $-1 \in \mathrm{SL}_2(\mathbb{Z})$, that the diagonal entries are positive. Finally, adding a suitable multiple of the second row to the first one, we obtain a matrix $\mathrm{UA} = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ with the required properties.

For the uniqueness, observe that a is the greatest common divisor of the elements in the first column of A (the operations performed to obtain an upper triangular form coincide with Euclid's algorithm). Now, $d = \frac{n}{a}$ and b is obviously uniquely determined modulo d. □

Lemma 3.18. *Let* $n \in \mathbb{N}$ *and let* $\mathrm{M}(n)$ *be the set of* 2×2 *matrices with integer entries and determinant* n. *Let* $\mathrm{X}(n)$ *be the subset of* $\mathrm{M}(n)$ *consisting of matrices of the form* $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ *with* $a \geq 1$ *and* $0 \leq b < d$. *If* $\Lambda = \Lambda(\omega_1, \omega_2)$, *then the sublattices of* Λ *of index* n *are precisely those of the form* $\Lambda(a\omega_1 + b\omega_2, d\omega_2)$ *for* $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \in \mathrm{X}(n)$.

Proof. If $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \in \mathrm{X}(n)$, then $\Lambda(a\omega_1 + b\omega_2, d\omega_2)$ has index n in Λ because $ad = n$. Conversely, if Λ' is a sublattice of Λ of index n, then every basis of Λ' must be of the form $(\omega_1', \omega_2') = (a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$ for some $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{M}(n)$. By lemma 3.17, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is $\mathrm{SL}_2(\mathbb{Z})$–equivalent to exactly one element of $\mathrm{X}(n)$. On the other hand, if there are two matrices $\alpha = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ and $\beta = \left(\begin{smallmatrix} a' & b' \\ 0 & d' \end{smallmatrix}\right)$ in $\mathrm{X}(n)$ giving rise to the same sublattice (i.e., such that $\Lambda(a\omega_1 + b\omega_2, d\omega_2) = \Lambda(a'\omega_1 + b'\omega_2, d'\omega_2)$), then $\alpha = u\beta$ for some $u \in \mathrm{SL}_2(\mathbb{Z})$ and so $\alpha = \beta$ (by lemma 3.17). □

Proposition 3.19. *Let n be a positive integer and let $\mathrm{M}(n)$ be the set of 2×2 matrices with integer entries and determinant n. Let $f : \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ be a weakly modular function for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$. We have that*

$$\mathrm{T}(n) f(z) = n^{2k-1} \sum_{a,b,d} d^{-2k} f\left(\frac{az+b}{d}\right) = n^{k-1} \sum_{\delta} f|_{2k}^{[\delta]}(z),$$

where the first sum is over the triples of integers a, b and d such that $a \geq 1$, $ad = n$ and $0 \leq b < d$ and the last sum is over a set of representatives of $\mathrm{SL}_2(\mathbb{Z}) \setminus \mathrm{M}(n)$.

Proof. The first equality is a direct consequence of lemma 3.18, while the second is a consequence of lemma 3.17. □

Corollary 3.20. *Let $f : \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ be a weakly modular function for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$. If f is holomorphic (resp. meromorphic) in $\mathbb{H}$, then $\mathrm{T}(n) f$ is also holomorphic (resp. meromorphic) in $\mathbb{H}$ for every $n \in \mathbb{N}$.*

Proposition 3.21. *Let $f : \mathbb{H}^* \to \mathbb{P}^1_{\mathbb{C}}$ be a meromorphic modular form (resp. modular form or cusp form) for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$ and consider its q–expansion at ∞*

$$\widehat{f_\infty}(q) = \sum_{m\in\mathbb{Z}} c(m) q^m.$$

For all $n \in \mathbb{N}$, the function $g = \mathrm{T}(n) f : \mathbb{H}^* \to \mathbb{P}^1_{\mathbb{C}}$ is also a meromorphic modular form (resp. modular form or cusp form) for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$ with q–expansion at ∞

$$\widehat{g_\infty}(q) = \sum_{m\in\mathbb{Z}} c_n(m) q^m$$

where, for all $m \in \mathbb{Z}$, the m–th coefficient is given by

$$c_n(m) = \sum_{a \mid (n,m)} a^{2k-1} c\left(\frac{mn}{a^2}\right)$$

(the last sum is over the positive divisors of (n, m)).

Proof. We already know that g is weakly modular of weight $2k$ and meromorphic (resp. holomorphic) in $\mathbb{H}$. Therefore, we need to prove that it is meromorphic (resp. holomorphic or zero) at ∞: this will be immediate from the form of the coefficients $c_n(m)$. We can write

$$\mathrm{T}(n) f(z) = n^{2k-1} \sum_{a\geq 1} \sum_{ad=n} \sum_{0\leq b<d} d^{-2k} \sum_{m\in\mathbb{Z}} c(m) e^{2\pi i m(az+b)/d}.$$

But, for fixed a and d, the sum $\sum_{0\le b<d} e^{2\pi i bm/d}$ is 0 unless $d \mid m$, in which case it is d. Thus, setting $m' = \frac{m}{d}$,

$$\mathrm{T}(n) f(z) = n^{2k-1} \sum_{a,d,m'} d^{-2k+1} c(m'd) e^{2\pi i a m' z}.$$

In the previous expression, we can collect powers of $e^{2\pi i z}$ to compute

$$c_n(t) = \sum_{a \mid (n,t)} a^{2k-1} c\left(\frac{n}{a}\frac{t}{a}\right)$$

(the sum is over the positive divisors of (n,t)). □

Corollary 3.22. *Let f be a non-zero modular form for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$ with q–expansion*

$$\widehat{f_\infty}(q) = \sum_{m=0}^{\infty} c_m q^m.$$

If f is an eigenform of all the $\mathrm{T}(n)$, with $\mathrm{T}(n) f = \lambda_n f$ for each $n \in \mathbb{N}$, then $c_1 \neq 0$ and $c_m = \lambda_m c_1$ for all $m \in \mathbb{N}$.

Proof. The coefficient of q in the q–expansion of $\mathrm{T}(n) f$ is c_n, by proposition 3.21. But, since $\mathrm{T}(n) f = \lambda_n f$, it is also $\lambda_n c_1$. Finally, if c_1 were zero, then c_m would be zero for all $m \in \mathbb{N}$, thus contradicting the assumption that f is non-zero. □

The previous result implies that the coefficients of the q–expansion of an eigenform (of all the $\mathrm{T}(n)$) satisfy the recurrence relations of proposition 3.16.

Lemma 3.23. *Let p be a prime and let $\mathrm{M}(p)$ be the set of 2×2 matrices with integer entries and determinant p. There exists a common set of representatives for the set of left orbits $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{M}(p)$ and for the set of right orbits $\mathrm{M}(p)/\mathrm{SL}_2(\mathbb{Z})$.*

Proof. Let $\alpha, \beta \in \mathrm{M}(p)$. Then,

$$\mathrm{SL}_2(\mathbb{Z})\alpha\,\mathrm{SL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})\beta\,\mathrm{SL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})\left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right)\mathrm{SL}_2(\mathbb{Z})$$

(this is the Smith normal form of every element of $\mathrm{M}(p)$). In particular, $\beta = u\alpha v$ for some $u, v \in \mathrm{SL}_2(\mathbb{Z})$. Therefore, we can take $\gamma = u\alpha = \beta v^{-1}$ and we obtain that $\mathrm{SL}_2(\mathbb{Z})\gamma = \mathrm{SL}_2(\mathbb{Z})\alpha$ and $\gamma\,\mathrm{SL}_2(\mathbb{Z}) = \beta\,\mathrm{SL}_2(\mathbb{Z})$.

Therefore, we can obtain representatives $\gamma_1, \ldots, \gamma_{p+1}$ for both $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{M}(p)$ and $\mathrm{M}(p)/\mathrm{SL}_2(\mathbb{Z})$ from a set of representatives $\alpha_1, \ldots, \alpha_{p+1}$ for $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{M}(p)$ and a set of representatives $\beta_1, \ldots, \beta_{p+1}$ for $\mathrm{M}(p)/\mathrm{SL}_2(\mathbb{Z})$. □

Theorem 3.24. *Let $f, g\colon \mathbb{H}^* \to \mathbb{P}^1_{\mathbb{C}}$ be two modular forms for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$ and suppose in addition that at least one of them is a cusp form. For all $n \in \mathbb{N}$, we have that $\langle \mathrm{T}(n) f, g\rangle = \langle f, \mathrm{T}(n) g\rangle$.*

Proof. Lemma 3.23 says that there exists a set of representatives $\gamma_1, \ldots, \gamma_{p+1}$ for both $\mathrm{SL}_2(\mathbb{Z}) \setminus \mathrm{M}(p)$ and $\mathrm{M}(p)/\mathrm{SL}_2(\mathbb{Z})$. For every j, write $\gamma_j' = p\gamma_j^{-1}$. We have that $\mathrm{M}(p) = \bigsqcup_j \mathrm{SL}_2(\mathbb{Z})\gamma_j = \bigsqcup_j \gamma_j \mathrm{SL}_2(\mathbb{Z})$ and so $\mathrm{M}(p) = p\mathrm{M}(p)^{-1} = \bigsqcup_j \mathrm{SL}_2(\mathbb{Z})\gamma_j'$ too.

Now, by proposition 3.19, we obtain that

$$\langle \mathrm{T}(p) f, g\rangle = p^{k-1} \sum_{j=1}^{p+1} \langle f|_{2k}^{[\gamma_j]}, g\rangle = p^{k-1} \sum_{j=1}^{p+1} \langle f, g|_{2k}^{[\gamma_j']}\rangle = \langle f, \mathrm{T}(p) g\rangle;$$

the general case follows from proposition 3.16. □

3.3 Hecke operators using double cosets

In the previous section, we described the action of Hecke operators on modular forms for the full modular group by interpreting them as functions on lattices. Our objective in this section is to generalise that construction and define operators $\mathrm{T}(n)$ on modular forms for congruence subgroups such as $\Gamma(\mathrm{N})$, $\Gamma_0(\mathrm{N})$ or $\Gamma_1(\mathrm{N})$ (for $\mathrm{N} \in \mathbb{N}$). These Hecke operators should satisfy essentially the same properties as Hecke operators for $\mathrm{SL}_2(\mathbb{Z})$.

Nevertheless, for a general congruence subgroup Γ, the set $\Gamma \setminus \mathbb{H}$ does not parametrise isomorphism classes of elliptic curves (or lattices modulo homothety) any more. Thus, one possible approach is to use lattices in conjunction with some additional torsion data (such as a subgroup of the lattice) in order to obtain a bijection with $\Gamma \setminus \mathbb{H}$. The downside of this strategy is that we should define what are known as modular points (lattices plus some additional structure) specifically for each kind of congruence subgroups we want to work with. The details of this method for $\Gamma_1(\mathrm{N})$ can be found, for instance, in the books [3] by Koblitz and [4] by Lang.

We follow a different approach which allows us to define Hecke operators directly on modular forms. Let f be a modular form for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$. Proposition 3.19 shows that $\mathrm{T}(n) f$ is a linear combination of terms of the form $f|_{2k}^{[\delta]}$ with $\delta \in \mathrm{GL}_2^+(\mathbb{Q})$ (in fact, $f|_{2k}^{[\delta]}$ depends only on the coset $\mathrm{SL}_2(\mathbb{Z})\delta$). Thus, if ω is the k–fold differential form on $\mathrm{X}(1)$ associated with f, then the k–fold differential form on $\mathrm{X}(1)$ associated with $\mathrm{T}(n) f$ is somehow related to the k–fold

differential forms $\delta^*(\omega)$ (which are not defined on $X(1)$ in general). We try to understand and generalise this situation.

Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. Let $\alpha \in GL_2^+(\mathbb{Q})$, which defines the map $z \mapsto \alpha(z) : \mathbb{H}^* \to \mathbb{H}^*$. We would like to define a map $\alpha\colon X(\Gamma) \to X(\Gamma)$. But $\alpha\Gamma z$ is not an element of $X(\Gamma)$ and $\Gamma\alpha(z)$ depends on the choice of the representative z. Therefore, we consider the union of the orbits meeting $\alpha\Gamma z$, which is $\Gamma\alpha\Gamma z$.

Lemma 3.25. *Let* $\Gamma' = \Gamma \cap \alpha^{-1}\Gamma\alpha$. *If* $\Gamma = \bigsqcup_{j=1}^{d} \Gamma'\beta_j$, *then* $\Gamma\alpha\Gamma = \bigsqcup_{j=1}^{d} \Gamma\alpha\beta_j$ *(which means, in particular, that* $\Gamma\alpha\Gamma$ *is the disjoint union of* $[\Gamma : \Gamma']$ *right cosets). Conversely, if* $\Gamma\alpha\Gamma = \bigsqcup_{j=1}^{d} \Gamma\alpha\beta_j$, *then* $\Gamma = \bigsqcup_{j=1}^{d} \Gamma'\beta_j$.

Proof. First, observe that Γ' is a congruence subgroup of $SL_2(\mathbb{Z})$ (in particular, of finite index), by lemma 3.7. This means that we can find a set of d representatives of $\Gamma' \backslash \Gamma$, indeed.

Suppose that $\Gamma = \bigsqcup_{j=1}^{d} \Gamma'\beta_j$. Consider an element $\gamma_1\alpha\gamma_2 \in \Gamma\alpha\Gamma$. We can write $\gamma_2 = \gamma'\beta_j$ with $\gamma' \in \Gamma'$ for some j. Since $\gamma' \in \alpha^{-1}\Gamma\alpha$, we can write $\gamma' = \alpha^{-1}\gamma\alpha$ for some $\gamma \in \Gamma$. In this case, we have that $\gamma_1\alpha\gamma_2 = \gamma_1\gamma\alpha\beta_j \in \Gamma\alpha\beta_j$. Therefore, $\Gamma\alpha\Gamma = \bigcup_{j=1}^{d} \Gamma\alpha\beta_j$: we must check that these cosets are disjoint.

Suppose that $\gamma_1\alpha\beta_i = \gamma_2\alpha\beta_j$ for some $\gamma_1, \gamma_2 \in \Gamma$ and some $i, j \in \{1, \dots, d\}$. We deduce that $\beta_i\beta_j^{-1} = \alpha^{-1}\gamma_1^{-1}\gamma_2\alpha \in \alpha^{-1}\Gamma\alpha$ and, since $\beta_i\beta_j^{-1} \in \Gamma$ as well, $\beta_i\beta_j^{-1} \in \Gamma'$, which is to say that $i = j$.

For the converse, assume that $\Gamma\alpha\Gamma = \bigsqcup_{j=1}^{d} \Gamma\alpha\beta_j$. Thus, if $\gamma \in \Gamma$, we can write $\alpha\gamma = \gamma_0\alpha\beta_j$ with $\gamma_0 \in \Gamma$ for some j. Consequently, $\gamma = \alpha^{-1}\gamma_0\alpha\beta_j \in \Gamma'\beta_j$. Indeed, this means that $\beta_j^{-1}\gamma = \alpha^{-1}\gamma_0\alpha$ and so this element belongs to both Γ and $\alpha^{-1}\Gamma\alpha$. In conclusion, $\Gamma = \bigcup_{j=1}^{d} \Gamma'\beta_j$: we must check that these cosets are disjoint.

Suppose that $\gamma_1'\beta_i = \gamma_2'\beta_j$ for some $\gamma_1', \gamma_2' \in \Gamma'$ and some $i, j \in \{1, \dots, d\}$. We can write $\gamma_1' = \alpha^{-1}\gamma_1\alpha$ and $\gamma_2' = \alpha^{-1}\gamma_2\alpha$ with $\gamma_1, \gamma_2 \in \Gamma$. Therefore, $\gamma_2^{-1}\gamma_1\alpha\beta_i = \alpha\beta_j$, which implies that $i = j$. □

Let $\Gamma_\alpha = \Gamma \cap \alpha^{-1}\Gamma\alpha$ and write $\Gamma = \bigsqcup_{j=1}^{d} \Gamma_\alpha\beta_j$. By lemma 3.25, the map induced by α (or by the double coset $\Gamma\alpha\Gamma$) on the modular curve $X(\Gamma)$ should be a "many-valued map" sending Γz to the d points $\Gamma\alpha\beta_j z$ (for $1 \le j \le d$). This idea can be formalised by means of the correspondence

$$\begin{array}{ccc} & X(\Gamma_\alpha) & \\ {\scriptstyle \iota}\swarrow & & \searrow{\scriptstyle \alpha} \\ X(\Gamma) & & X(\Gamma) \end{array}$$

(where $\alpha(\Gamma_\alpha z) = \Gamma\alpha(z)$ and $\iota(\Gamma_\alpha z) = \Gamma z$). It is called the Hecke correspondence because the action of Hecke operators on k–fold differential forms on $X(\Gamma)$ can be defined (using appropriate matrices α) as the pull-back by α followed by the trace given by ι in the above diagram. We could also interpret this correspondence as a map on divisors given by $\Gamma z \mapsto \sum_{j=1}^{d} \Gamma(\alpha\beta_j)(z) : \mathrm{Div}(X(\Gamma)) \to \mathrm{Div}(X(\Gamma))$.

The previous discussion leads us to define an action of double cosets on modular forms. In the remainder of this chapter, k will be an integer.

Definition 3.26. Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. We define a right action of weight k of double cosets of the form $\Gamma\alpha\Gamma$ with $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ on weakly modular functions $f : \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ for Γ of weight k in the following way:

$$f|_k^{[\Gamma\alpha\Gamma]}(z) = \sum_{j=1}^{d} f|_k^{[\alpha_j]}(z),$$

where $\Gamma\alpha\Gamma = \bigsqcup_{j=1}^{d} \Gamma\alpha_j$.

This action is well-defined. Firstly, if α_j is replaced by $\gamma_j\alpha_j$ with $\gamma_j \in \Gamma$, then $f|_k^{[\gamma_j\alpha_j]} = f|_k^{[\alpha_j]}$ because f is weakly modular for Γ of weight k. Secondly, if α is replaced by α' such that $\Gamma\alpha\Gamma = \Gamma\alpha'\Gamma$, we can still take the same decomposition as a disjoint union of right cosets: $\Gamma\alpha'\Gamma = \bigsqcup_{j=1}^{d} \Gamma\alpha_j$. Finally, by lemma 3.25, we can take $\alpha_j = \alpha\beta_j$ for all j so that $\Gamma = \bigsqcup_{j=1}^{d} \Gamma_\alpha\beta_j$. Let $\gamma \in \Gamma$. Since multiplication by γ permutes the right cosets $\Gamma_\alpha\beta_j$, we conclude that

$$\left(f|_k^{[\Gamma\alpha\Gamma]}\right)\Big|_k^{[\gamma]} = \sum_{j=1}^{d} f|_k^{[\alpha_j\gamma]} = \sum_{j=1}^{d} f|_k^{[\alpha_j]} = f|_k^{[\Gamma\alpha\Gamma]}$$

and so $f|_k^{[\Gamma\alpha\Gamma]}$ is weakly modular for Γ of weight k.

Proposition 3.27. *Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. If $f \in M_k(\Gamma)$, then $f|_k^{[\Gamma\alpha\Gamma]} \in M_k(\Gamma)$. Furthermore, if $f \in S_k(\Gamma)$, then $f|_k^{[\Gamma\alpha\Gamma]} \in S_k(\Gamma)$.*

Proof. We already know that $f|_k^{[\Gamma\alpha\Gamma]}$ is weakly modular for Γ of weight k. Furthermore, $f|_k^{[\Gamma\alpha\Gamma]}$ is holomorphic in $\mathbb{H}$ and satisfies the required conditions at the cusps because each $f|_k^{[\alpha_j]}$ appearing in its definition does (by lemma 3.7). □

Definition 3.28. Let N be a positive integer. Let S^+ be a non-trivial subgroup of $\mathbb{Z}$ (i.e., $S^+ = M\mathbb{Z}$ for some $M \in \mathbb{N}$) and let $S^\times$ be a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$ (which

we identify with its inverse image in $\mathbb{Z}$ under the projection modulo N). Let n be a positive integer. We define

$$\Delta^n(\mathrm{N},\mathrm{S}^\times,\mathrm{S}^+)=\left\{\begin{pmatrix}a & b\\ c & d\end{pmatrix}: a,b,c,d\in\mathbb{Z},\ a\in\mathrm{S}^\times,\ b\in\mathrm{S}^+,\ \mathrm{N}\,\middle|\, c \text{ and } ad-bc=n\right\}.$$

Example 3.29. If $\mathrm{N}=1$ and $\mathrm{S}^\times=\mathrm{S}^+=\mathbb{Z}$, then $\Delta^n(\mathrm{N},\mathrm{S}^\times,\mathrm{S}^+)$ is the set of 2×2 matrices with integer entries and determinant n.

In general, one checks easily that

$$\Delta^m(\mathrm{N},\mathrm{S}^\times,\mathrm{S}^+)\cdot\Delta^n(\mathrm{N},\mathrm{S}^\times,\mathrm{S}^+)\subseteq\Delta^{mn}(\mathrm{N},\mathrm{S}^\times,\mathrm{S}^+);$$

moreover, $\Delta^1(\mathrm{N},\mathrm{S}^\times,\mathrm{S}^+)$ is a group (actually, a congruence subgroup, since it contains $\Gamma([\mathrm{M},\mathrm{N}])$ if $\mathrm{S}^+=\mathrm{M}\mathbb{Z}$). In particular, we can express $\Gamma(\mathrm{N})=\Delta^1(\mathrm{N},\{1\},\mathrm{N}\mathbb{Z})$, $\Gamma_0(\mathrm{N})=\Delta^1(\mathrm{N},(\mathbb{Z}/\mathrm{N}\mathbb{Z})^\times,\mathbb{Z})$ and $\Gamma_1(\mathrm{N})=\Delta^1(\mathrm{N},\{1\},\mathbb{Z})$.

Definition 3.30. Let $\Gamma=\Delta^1(\mathrm{N},\mathrm{S}^\times,\mathrm{S}^+)$ (with the notation of definition 3.28) and let $n\in\mathbb{N}$. We define the action of the n–th *Hecke operator* on modular forms for Γ of weight k, $\mathrm{T}(n)\colon \mathrm{M}_k(\Gamma)\to\mathrm{M}_k(\Gamma)$, as follows: for all $f\in\mathrm{M}_k(\Gamma)$,

$$\mathrm{T}(n)f=n^{\frac{k}{2}-1}\sum f|_k^{[\Gamma\alpha\Gamma]}=n^{\frac{k}{2}-1}\sum f|_k^{[\beta]},$$

where the first sum is over all the double cosets $\Gamma\alpha\Gamma$ contained in $\Delta^n(\mathrm{N},\mathrm{S}^\times,\mathrm{S}^+)$ and the second sum is over the elements $\Gamma\beta$ of $\Gamma\backslash\Delta^n(\mathrm{N},\mathrm{S}^\times,\mathrm{S}^+)$.

Now that we have defined Hecke operators for a large class of congruence subgroups, we focus on the groups $\Gamma_0(\mathrm{N})$ for $\mathrm{N}\in\mathbb{N}$. Fix $\mathrm{N}\in\mathbb{N}$. The proofs of the following results are adapted from Miyake's book [9] and Koblitz's book [3] (which deals with the case of $\Gamma_1(\mathrm{N})$).

Lemma 3.31. *We can express*

$$\Delta^n(\mathrm{N},(\mathbb{Z}/\mathrm{N}\mathbb{Z})^\times,\mathbb{Z})=\bigsqcup_{a,b,d}\Gamma_0(\mathrm{N})\begin{pmatrix}a & b\\ 0 & d\end{pmatrix},$$

where the (disjoint) union is over the triples of integers a, b and d such that $a\geq 1$, $(a,\mathrm{N})=1$, $ad=n$ and $0\leq b<d$. In addition,

$$\Delta^n(\mathrm{N},(\mathbb{Z}/\mathrm{N}\mathbb{Z})^\times,\mathbb{Z})=\bigsqcup_{a,d}\Gamma_0(\mathrm{N})\begin{pmatrix}a & 0\\ 0 & d\end{pmatrix}\Gamma_0(\mathrm{N}),$$

where the (disjoint) union is over the pairs of integers a and d such that $a \geq 1$, $(a, \mathrm{N}) = 1$, $ad = n$ and $a \mid d$.

Proof. Let $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Delta^n(\mathrm{N}, (\mathbb{Z}/\mathrm{N}\mathbb{Z})^\times, \mathbb{Z})$. Let $\mathrm{D} = (a, c)$ and consider $x, y \in \mathbb{Z}$ such that $ax + cy = \mathrm{D}$ (Bézout's identity). Write $\mathrm{A} = \frac{a}{\mathrm{D}}$ and $\mathrm{C} = \frac{c}{\mathrm{D}}$, so that $\mathrm{A}x + \mathrm{C}y = 1$. Since $(a, \mathrm{N}) = 1$ and $\mathrm{N} \mid c$, we obtain that $\mathrm{N} \mid \mathrm{C}$ and $(\mathrm{D}, \mathrm{N}) = 1$. As a consequence, $\gamma = \left(\begin{smallmatrix} x & y \\ -\mathrm{C} & \mathrm{A} \end{smallmatrix}\right) \in \Gamma_0(\mathrm{N})$. Now observe that $\gamma\alpha = \left(\begin{smallmatrix} \mathrm{D} & bx+dy \\ 0 & -b\mathrm{C}+d\mathrm{A} \end{smallmatrix}\right)$. Hence, $\pm\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right)\gamma\alpha$ is of the form $\left(\begin{smallmatrix} a' & b' \\ 0 & d' \end{smallmatrix}\right)$ with $a' \geq 1$, $(a', \mathrm{N}) = 1$, $a'd' = n$ and $0 \leq b' < d'$ (for a suitable choice of h).

Moreover, for any two matrices $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} a' & b' \\ 0 & d' \end{smallmatrix}\right)$ of the desired form, we have that $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)^{-1}\left(\begin{smallmatrix} a' & b' \\ 0 & d' \end{smallmatrix}\right) = \left(\begin{smallmatrix} a^{-1}a' & * \\ 0 & d^{-1}d' \end{smallmatrix}\right)$ and this matrix is in $\Gamma_0(\mathrm{N})$ if and only if $a = a'$, $d = d'$ and $b = b'$.

The proof of the first statement is complete. The second statement is analogous to the existence and uniqueness of a Smith normal form.

Let $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Delta^n(\mathrm{N}, (\mathbb{Z}/\mathrm{N}\mathbb{Z})^\times, \mathbb{Z})$. We have already seen that we can transform α (multiplying on the left by a matrix of $\Gamma_0(\mathrm{N})$) into a matrix of the form $\left(\begin{smallmatrix} a' & b' \\ 0 & d' \end{smallmatrix}\right)$ with $|a'| < |a|$ if $c \neq 0$. Similarly, we can transform it (multiplying on the right by a matrix of $\Gamma_0(\mathrm{N})$) into a matrix of the form $\left(\begin{smallmatrix} a' & 0 \\ c' & d' \end{smallmatrix}\right)$ with $|a'| < |a|$ if $b \neq 0$. Indeed, let $\mathrm{D} = (a, b)$ and write $\mathrm{A} = \frac{a}{\mathrm{D}}$ and $\mathrm{B} = \frac{b}{\mathrm{D}}$. Since $(a, \mathrm{N}) = 1$, we obtain that $(\mathrm{D}, \mathrm{N}) = (\mathrm{A}, \mathrm{N}) = 1$. Take $x, y \in \mathbb{Z}$ such that $\mathrm{A}x + \mathrm{B}y = 1$ (Bézout's identity) and $\mathrm{N} \mid y$ (we can do so because $(\mathrm{A}, \mathrm{N}) = 1$ and the possible choices of y differ by multiples of A). Therefore, $\gamma = \left(\begin{smallmatrix} x & \mathrm{B} \\ y & -\mathrm{A} \end{smallmatrix}\right) \in \Gamma_0(\mathrm{N})$ and $\alpha\gamma = \left(\begin{smallmatrix} \mathrm{D} & 0 \\ cx+dy & c\mathrm{B}-d\mathrm{A} \end{smallmatrix}\right)$.

Alternating between these two types of transformations, we can transform any matrix $\alpha \in \Delta^n(\mathrm{N}, (\mathbb{Z}/\mathrm{N}\mathbb{Z})^\times, \mathbb{Z})$ into a matrix of the form $\left(\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right)$ (because the absolute value of the first entry of the matrix decreases at each step). Now, multiplying on the right by $\left(\begin{smallmatrix} 1 & 0 \\ \mathrm{N} & 1 \end{smallmatrix}\right)$, we obtain the matrix $\left(\begin{smallmatrix} a & 0 \\ \mathrm{N}d & d \end{smallmatrix}\right)$. This matrix turns into $\left(\begin{smallmatrix} (a,d) & dy \\ 0 & d\mathrm{A} \end{smallmatrix}\right)$ with a transformation of the first kind (observe that $(a, d) = (a, \mathrm{N}d)$ because $(a, \mathrm{N}) = 1$) and then into $\left(\begin{smallmatrix} (a,d) & 0 \\ 0 & d\mathrm{A} \end{smallmatrix}\right)$ with a transformation of the second kind. And it is clear that $(a, d) \mid d\mathrm{A}$.

The fact that these double cosets are disjoint follows from the uniqueness of the Smith normal form of a matrix with integer coefficients. □

Proposition 3.32. *Let $f : \mathbb{H}^* \to \mathbb{P}^1_{\mathbb{C}}$ be a meromorphic modular form (resp. modular form or cusp form) for $\Gamma_0(\mathrm{N})$ of weight k and consider its q–expansion at ∞*

$$\widehat{f_\infty}(q) = \sum_{m \in \mathbb{Z}} c(m) q^m .$$

For all $n \in \mathbb{N}$, the function $g = \mathrm{T}(n) f : \mathbb{H}^* \to \mathbb{P}^1_{\mathbb{C}}$ is also a meromorphic modular form (resp. modular form or cusp form) for $\Gamma_0(\mathrm{N})$ of weight k with q–expansion at ∞

$$\widehat{g_\infty}(q) = \sum_{m \in \mathbb{Z}} c_n(m) q^m$$

where, for all $m \in \mathbb{Z}$, the m–th coefficient is given by

$$c_n(m) = \sum_{a \mid (n,m)} \chi_{\mathrm{N}}(a) a^{k-1} c\left(\frac{mn}{a^2}\right)$$

(the last sum is over the positive divisors of (n, m) which are prime to N: here, χ_{N} denotes the principal Dirichlet character modulo N).

Proof. This result is analogous to proposition 3.21. By lemma 3.31, we can write

$$\mathrm{T}(n) f(z) = n^{k-1} \sum_{(a,\mathrm{N})=1} \sum_{ad=n} \sum_{0 \le b < d} d^{-k} \sum_{m \in \mathbb{Z}} c(m) e^{2\pi i m(az+b)/d}$$

(a is positive in this sum). Since, for fixed a and d, the sum $\sum_{0 \le b < d} e^{2\pi i b m/d}$ is 0 unless $d \mid m$, we set $m' = \frac{m}{d}$ and

$$\mathrm{T}(n) f(z) = n^{k-1} \sum_{a,d,m'} d^{-k+1} c(m'd) e^{2\pi i a m' z}.$$

Collecting powers of $e^{2\pi i z}$ in the previous expression, we obtain that

$$c_n(t) = \sum_{\substack{(a,\mathrm{N})=1 \\ a \mid (n,t)}} a^{k-1} c\left(\frac{nt}{a^2}\right) = \sum_{a \mid (n,t)} \chi_{\mathrm{N}}(a) a^{k-1} c\left(\frac{nt}{a^2}\right)$$

(the sum is over the positive divisors of (n, t) which are prime to N). □

The previous results allow us to compute explicitly the action of Hecke operators on modular forms for $\Gamma_0(\mathrm{N})$. Next, we are going to prove that there is a basis of $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ whose elements are eigenforms of many of the $\mathrm{T}(n)$.

Lemma 3.33. *Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. There exists a common set of representatives for $\Gamma \backslash \Gamma\alpha\Gamma$ and for $\Gamma\alpha\Gamma/\Gamma$.*

Proof. Write $\Gamma\alpha\Gamma = \bigsqcup_{j=1}^{d} \Gamma\alpha_j = \bigsqcup_{j=1}^{d} \beta_j \Gamma$ (observe that the number of left cosets and the number of right cosets coincide: it is $d = [\Gamma : (\Gamma \cap \alpha^{-1}\Gamma\alpha)]$, by lemma 3.25).

For $i, j \in \{1, \dots, d\}$, $\Gamma\alpha_i \cap \beta_j\Gamma \neq \emptyset$: if that were not the case, we would have that $\Gamma\alpha\Gamma = \Gamma\alpha_i\Gamma \subseteq \bigcup_{l\neq j} \beta_l\Gamma$, but this is impossible. Thus, we can choose an element $\delta_j \in \Gamma\alpha_j \cap \beta_j\Gamma$ for each $j \in \{1, \dots, d\}$. Since $\Gamma\alpha_j = \Gamma\delta_j$ and $\beta_j\Gamma = \delta_j\Gamma$ for every j, we conclude that $\Gamma\alpha\Gamma = \bigsqcup_{j=1}^{d} \Gamma\delta_j = \bigsqcup_{j=1}^{d} \delta_j\Gamma$. □

Lemma 3.34. *With the notation of definition 3.28, let* $\Gamma = \Delta^1(\mathrm{N}, \mathrm{S}^\times, \mathrm{S}^+)$ *and consider* $\Delta = \bigsqcup_{n\in\mathbb{N}} \Delta^n(\mathrm{N}, \mathrm{S}^\times, \mathrm{S}^+)$. *Assume that there exists a map* $\alpha \mapsto \alpha^\iota : \Delta \to \Delta$ *satisfying that*

(i) $(\alpha\beta)^\iota = \beta^\iota\alpha^\iota$ *and* $(\alpha^\iota)^\iota = \alpha$ *for all* $\alpha, \beta \in \Delta$,

(ii) $\Gamma^\iota = \Gamma$, *and*

(iii) $\Gamma\alpha^\iota\Gamma = \Gamma\alpha\Gamma$ *for all* $\alpha \in \Delta$.

Then, for every weakly modular form $f : \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ *for* Γ *of weight* k *and for all* $m, n \in \mathbb{N}$, $\mathrm{T}(m)\mathrm{T}(n)f = \mathrm{T}(n)\mathrm{T}(m)f$.

Proof. Let $\alpha \in \Delta^m(\mathrm{N}, \mathrm{S}^\times, \mathrm{S}^+)$ and $\beta \in \Delta^n(\mathrm{N}, \mathrm{S}^\times, \mathrm{S}^+)$. By lemma 3.33, we can express $\Gamma\alpha\Gamma = \bigsqcup_{i=1}^{a} \Gamma\alpha_i = \bigsqcup_{i=1}^{a} \alpha_i\Gamma$ and $\Gamma\beta\Gamma = \bigsqcup_{j=1}^{b} \Gamma\beta_j = \bigsqcup_{j=1}^{b} \beta_j\Gamma$. Thus, the properties of ι imply that $\Gamma\alpha\Gamma = \Gamma\alpha^\iota\Gamma = (\Gamma\alpha\Gamma)^\iota = \bigsqcup_{i=1}^{a} \alpha_i^\iota\Gamma = \bigsqcup_{i=1}^{a} \Gamma\alpha_i^\iota$ and, analogously, $\Gamma\beta\Gamma = \bigsqcup_{j=1}^{b} \beta_j^\iota\Gamma = \bigsqcup_{j=1}^{d} \Gamma\beta_j^\iota$. Therefore,

$$\left(f|_k^{[\Gamma\alpha\Gamma]}\right)\Big|_k^{[\Gamma\beta\Gamma]} = \sum_{j=1}^{b}\sum_{i=1}^{a} f|_k^{[\alpha_i\beta_j]} \quad \text{and} \quad \left(f|_k^{[\Gamma\beta\Gamma]}\right)\Big|_k^{[\Gamma\alpha\Gamma]} = \sum_{i=1}^{a}\sum_{j=1}^{b} f|_k^{[\beta_j^\iota\alpha_i^\iota]};$$

we want to check that each coset $\Gamma\delta$ appears the same number of times in each of these expressions. Indeed, $\alpha_i\beta_j \in \Gamma\delta\Gamma$ if and only if $\alpha_i\beta_j \in \Gamma\delta\gamma$ for some $\gamma \in \Gamma$ (and there are $|\Gamma \setminus \Gamma\delta\Gamma|$ cosets of the form $\Gamma\delta\gamma$). Hence,

$$\begin{aligned} c(\delta) = \left|\{(i,j) : \Gamma\alpha_i\beta_j = \Gamma\delta\}\right| &= \frac{\left|\{(i,j) : \Gamma\alpha_i\beta_j\Gamma = \Gamma\delta\Gamma\}\right|}{|\Gamma \setminus \Gamma\alpha\Gamma|} = \frac{\left|\{(i,j) : \Gamma\beta_j^\iota\alpha_i^\iota\Gamma = \Gamma\delta^\iota\Gamma\}\right|}{|\Gamma \setminus \Gamma\delta^\iota\Gamma|} \\ &= \frac{\left|\{(i,j) : \Gamma\beta_j^\iota\alpha_i^\iota\Gamma = \Gamma\delta\Gamma\}\right|}{|\Gamma \setminus \Gamma\delta\Gamma|} = \left|\{(i,j) : \Gamma\beta_j^\iota\alpha_i^\iota = \Gamma\delta\}\right| = c'(\delta) \end{aligned}$$

(here, $c(\delta)$ is the number of times $\Gamma\delta$ appears in the first expression and $c'(\delta)$ is the number of times $\Gamma\delta$ appears in the second expression).

Since the action of Hecke operators is defined in terms of $\mathbb{Z}$–linear combinations of double cosets and we have proved that the actions of $\Gamma\alpha\Gamma$ and of $\Gamma\beta\Gamma$ commute, we conclude that $\mathrm{T}(m)\mathrm{T}(n)f = \mathrm{T}(n)\mathrm{T}(m)f$. □

Proposition 3.35. *Let* $f : \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ *be a weakly modular form for* $\Gamma_0(\mathrm{N})$ *of weight* k. *For all* $m, n \in \mathbb{N}$, $\mathrm{T}(m)\mathrm{T}(n)f = \mathrm{T}(n)\mathrm{T}(m)f$.

Proof. Let $\Delta = \bigsqcup_{n\in\mathbb{N}} \Delta^n(\mathrm{N}, (\mathbb{Z}/\mathrm{N}\mathbb{Z})^\times, \mathbb{Z})$. Observe that the map

$$\begin{aligned} \iota\colon \Delta &\longrightarrow \Delta \\ \begin{pmatrix} a & b \\ \mathrm{N}c & d \end{pmatrix} &\longmapsto \begin{pmatrix} a & c \\ \mathrm{N}b & d \end{pmatrix} \end{aligned}$$

satisfies the three conditions of lemma 3.34. Indeed, conditions (i) and (ii) are obvious by definition, and condition (iii) is immediate choosing diagonal matrices as representatives of the double cosets (as in lemma 3.31). Therefore, the proposition is a consequence of lemma 3.34. □

In fact, a stronger result holds. We state it without proof.

Proposition 3.36. *Let $f : \mathbb{H} \to \mathbb{P}^1_{\mathbb{C}}$ be a weakly modular form for $\Gamma_0(\mathrm{N})$ of weight k. Then:*

(1) $\mathrm{T}(m)\mathrm{T}(n)f = \mathrm{T}(mn)f$ *for all $m, n \in \mathbb{N}$ such that $(m, n) = 1$;*
(2) $\mathrm{T}(p)\mathrm{T}(p^n)f = \mathrm{T}(p^{n+1})f + p^{k-1}\mathrm{T}(p^{n-1})f$ *for all prime p such that $p \nmid \mathrm{N}$ and all $n \in \mathbb{N}$;*
(3) $\mathrm{T}(p)\mathrm{T}(p^n)f = \mathrm{T}(p^{n+1})f$ *for all prime p such that $p \mid \mathrm{N}$ and all $n \in \mathbb{N}$.*

Theorem 3.37. *Let $f, g \in \mathbb{H}^* \to \mathbb{P}^1_{\mathbb{C}}$ be two modular forms for $\Gamma_0(\mathrm{N})$ of weight k and suppose in addition that at least one of them is a cusp form. For all $n \in \mathbb{N}$ such that $(n, \mathrm{N}) = 1$, $\langle \mathrm{T}(n)f, g\rangle = \langle f, \mathrm{T}(n)g\rangle$.*

Proof. For every $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, write $\alpha' = \det(\alpha)\alpha^{-1}$. Proposition 3.8 implies that $\langle f|_k^{[\Gamma_0(\mathrm{N})\alpha\Gamma_0(\mathrm{N})]}, g\rangle = \langle f, g|_k^{[\Gamma_0(\mathrm{N})\alpha'\Gamma_0(\mathrm{N})]}\rangle$.

Consider the set $\mathrm{X}(n) = \left\{ \left(\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right) : ad = n \text{ and } a \mid d \right\}$. By lemma 3.31, $\mathrm{X}(n)$ is a set of representatives of the double cosets $\Gamma_0(\mathrm{N})\alpha\Gamma_0(\mathrm{N})$ in $\Delta^n(\mathrm{N}, (\mathbb{Z}/\mathrm{N}\mathbb{Z})^\times, \mathbb{Z})$. If $\alpha = \left(\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right) \in \mathrm{X}(n)$, then $\alpha' = \left(\begin{smallmatrix} d & 0 \\ 0 & a \end{smallmatrix}\right) \in \Delta^n(\mathrm{N}, (\mathbb{Z}/\mathrm{N}\mathbb{Z})^\times, \mathbb{Z})$ because $(n, \mathrm{N}) = 1$ and $ad = n$. Moreover, the last part of the proof of lemma 3.31 shows that $\left(\begin{smallmatrix} d & 0 \\ 0 & a \end{smallmatrix}\right)$ can be transformed into $\left(\begin{smallmatrix} (a,d) & 0 \\ 0 & ad/(a,d) \end{smallmatrix}\right) = \left(\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right)$ with operations which correspond to multiplication by matrices of $\Gamma_0(\mathrm{N})$. That is, $\Gamma_0(\mathrm{N})\alpha'\Gamma_0(\mathrm{N}) = \Gamma_0(\mathrm{N})\alpha\Gamma_0(\mathrm{N})$.

In conclusion,

$$\begin{aligned} \langle \mathrm{T}(n)f, g\rangle &= n^{\frac{k}{2}-1} \sum_{\alpha\in\mathrm{X}(n)} \langle f|_k^{[\Gamma_0(\mathrm{N})\alpha\Gamma_0(\mathrm{N})]}, g\rangle = n^{\frac{k}{2}-1} \sum_{\alpha\in\mathrm{X}(n)} \langle f, g|_k^{[\Gamma_0(\mathrm{N})\alpha'\Gamma_0(\mathrm{N})]}\rangle \\ &= n^{\frac{k}{2}-1} \sum_{\alpha\in\mathrm{X}(n)} \langle f, g|_k^{[\Gamma_0(\mathrm{N})\alpha\Gamma_0(\mathrm{N})]}\rangle = \langle f, \mathrm{T}(n)g\rangle, \end{aligned}$$

as claimed. □

Corollary 3.38. *For every $k \in \mathbb{Z}$, there exists a basis of the complex vector space $S_k(\Gamma_0(N))$ whose elements are eigenforms of all the $T(n)$ for $n \in \mathbb{N}$ with $(n, N) = 1$.*

We could also be interested in finding eigenforms of the $T(n)$ with $(n, N) > 1$. If the eigenspaces of the $T(n)$ with $(n, N) > 1$ had dimension 1, then there would be a basis of eigenforms for all the Hecke operators (because Hecke operators commute and so preserve each eigenspace). However, this is not true in general. To overcome this barrier, one must study which modular forms for $\Gamma_0(N)$ come from lower levels. That is, if $N = d_1 d_2$ and $f \in M_k(\Gamma_0(d_1))$, then $f(z) \in M_k(\Gamma_0(N))$ and also $g(z) = f(d_2 z) \in M_k(\Gamma_0(N))$. The subspace of $S_k(\Gamma_0(N))$ spanned by the forms obtained in these two ways from elements of $S_k(\Gamma_0(d))$ for proper divisors d of N is called the subspace of oldforms. Its orthogonal complement is called the subspace of newforms. It can be shown that the spaces of newforms have bases composed of eigenforms of all the $T(n)$ (for $n \in \mathbb{N}$).

Another question is whether there is a basis of eigenforms of $M_k(\Gamma_0(N))$. In general, we cannot use the Petersson inner product (because at least one of the two forms must be a cusp form in order to ensure convergence). Instead, one can define explicitly generalised Eisenstein series which are eigenforms. In this case, $M_k(\Gamma_0(N))$ can be decomposed as the orthogonal direct sum of $S_k(\Gamma_0(N))$ and the space spanned by these generalised Eisenstein series.

The theories of oldforms and newforms and of generalised Eisenstein series are introduced, for instance, in the book [2] by Diamond and Shurman.

Chapter 4

Modular symbols

We are finally in a position to describe explicitly a basis of cusp forms by means of the theory of modular symbols. In particular, modular symbols provide a method of computing the Fourier expansions of the elements of a basis of cusp forms for congruence subgroups. Furthermore, the theory of modular symbols is an important tool in some proofs of theoretical results. As a matter of fact, the many applications of this theory make it a valuable tool in its own right.

In this chapter, we focus on the study of $S_2(\Gamma_0(N))$ for any $N \in \mathbb{N}$. This is the simplest case, yet it exhibits the most important aspects of this theory. The generalisation to cusp forms of weight k greater than 2 is based on the same ideas; however, it involves the use of complex polynomials in two variables which are homogeneous of degree $k-2$, which makes the proofs more complicated.

The main ideas of this theory, with a special emphasis on computations, are explained in the books [14] by Stein and [1] by Cremona. In a similar fashion, Stein's paper [13] summarises the most important results from a computational viewpoint and Merel's paper [6] finds simple sets of matrices which are sufficient to perform all the computations. Also, Lang's book [4] explains some results related to modular symbols and some of their theoretical applications. Nevertheless, all these references skip many proofs. Thus, most of the proofs in this chapter are adapted directly from the first half of Manin's original paper [5].

4.1 Motivation

Throughout this chapter, let N be a positive integer. We are finally going to take advantage of the theory explained in the previous chapters in order to find a basis of $S_2(\Gamma_0(N))$.

Fix $k \in \mathbb{N}$ (in fact, we are going to be interested in the case $k = 2$). Corollary 3.38 ensures the existence of a basis of $S_k(\Gamma_0(N))$ formed of eigenforms of all the $T(n)$ with $(n, N) = 1$. Hence, the action of Hecke operators should give us a fair amount of information about $S_k(\Gamma_0(N))$.

Let $n \in \mathbb{N}$. By proposition 3.32, $\mathrm{T}(n)$ acts on the q–expansions (at infinity) of modular forms for $\Gamma_0(\mathrm{N})$ of weight k in the following way:

$$\mathrm{T}(n)\Big(\sum_{m=0}^{\infty} a_m q^m\Big) = \sum_{m=0}^{\infty} \sum_{d \mid (n,m)} \chi_{\mathrm{N}}(d) d^{k-1} a_{mn/d^2} q^m,$$

where χ_{N} is the principal Dirichlet character modulo N and the last sum is only over positive divisors d of (n, m). For every $f \in \mathrm{M}_k(\Gamma_0(\mathrm{N}))$ and all $m \in \mathbb{N}$, we write $a_m(f)$ for the coefficient of q^m in $\widehat{f_\infty}(q)$.

Lemma 4.1. *Let $n \in \mathbb{N}$. For all $f \in \mathrm{M}_k(\Gamma_0(\mathrm{N}))$, $a_1(\mathrm{T}(n) f) = a_n(f)$.*

Let $\mathbb{T}$ be the $\mathbb{Z}$–algebra generated by the operators $\mathrm{T}(n)$ for $n \in \mathbb{N}$ acting on $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ (regarded as a subalgebra of $\mathrm{End}(\mathrm{S}_k(\Gamma_0(\mathrm{N})))$ and consider the complex vector space $\mathbb{T}_{\mathbb{C}} = \mathbb{T} \otimes_{\mathbb{Z}} \mathbb{C}$ obtained by extension of scalars.

Proposition 4.2. *There is a perfect pairing of complex vector spaces given by*

$$\begin{aligned} \langle \cdot, \cdot \rangle \colon \mathrm{S}_k(\Gamma_0(\mathrm{N})) \times \mathbb{T}_{\mathbb{C}} &\longrightarrow \mathbb{C} \\ (f, \mathrm{T}) &\longmapsto \langle f, \mathrm{T} \rangle = a_1(\mathrm{T} f) \end{aligned}$$

and, thus, it defines an isomorphism between $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ and $\mathrm{Hom}_{\mathbb{C}}(\mathbb{T}_{\mathbb{C}}, \mathbb{C})$.

Proof. This pairing is bilinear because every $\mathrm{T} \in \mathbb{T}_{\mathbb{C}}$ belongs to $\mathrm{End}_{\mathbb{C}}(\mathrm{S}_k(\Gamma_0(\mathrm{N})))$ and a_1 is also a linear map.

Let $f \in \mathrm{S}_k(\Gamma_0(\mathrm{N}))$ such that $\langle f, \mathrm{T} \rangle = 0$ for all $\mathrm{T} \in \mathbb{T}_{\mathbb{C}}$. By lemma 4.1, we have that $a_n(f) = \langle f, \mathrm{T}(n) \rangle = 0$ for all $n \in \mathbb{N}$, which means that $f = 0$.

Similarly, let $\mathrm{T} \in \mathbb{T}_{\mathbb{C}}$ such that $\langle f, \mathrm{T} \rangle = 0$ for all $f \in \mathrm{S}_k(\Gamma_0(\mathrm{N}))$. Fix $f \in \mathrm{S}_k(\Gamma_0(\mathrm{N}))$. By proposition 3.35, $\mathbb{T}_{\mathbb{C}}$ is commutative. In particular, for all $n \in \mathbb{N}$,

$$a_n(\mathrm{T} f) = a_1(\mathrm{T}(n)\,\mathrm{T} f) = a_1(\mathrm{T}\,\mathrm{T}(n) f) = \langle \mathrm{T}(n) f, \mathrm{T} \rangle = 0.$$

That is, $\mathrm{T} f = 0$. Since f was arbitrary, we conclude that T is the 0 operator.

Finally, considering that $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ has finite dimension, the pairing must be perfect and so induces an isomorphism between $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ and $\mathrm{Hom}_{\mathbb{C}}(\mathbb{T}_{\mathbb{C}}, \mathbb{C})$. □

Proposition 4.3. *Consider the isomorphism of complex vector spaces*

$$\begin{aligned} \Psi \colon \mathrm{S}_k(\Gamma_0(\mathrm{N})) &\longrightarrow \mathrm{Hom}_{\mathbb{C}}(\mathbb{T}_{\mathbb{C}}, \mathbb{C}) \\ f &\longmapsto \big(\mathrm{T} \mapsto a_1(\mathrm{T} f)\big) \end{aligned}$$

induced by the perfect pairing described in proposition 4.2. For every $\mathbb{C}$–linear map $\varphi\colon \mathbb{T}_{\mathbb{C}} \to \mathbb{C}$, the power series

$$f_\varphi(q) = \sum_{n=1}^{\infty} \varphi(\mathrm{T}(n))q^n$$

is the q–expansion at infinity of the cusp form $\Psi^{-1}(\varphi)$.

Proof. Let $g = \Psi^{-1}(\varphi)$. By definition, g is the only element of $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ such that $\langle g, \mathrm{T}\rangle = \varphi(\mathrm{T})$ for all $\mathrm{T} \in \mathbb{T}_{\mathbb{C}}$. In particular, by lemma 4.1,

$$a_n(g) = a_1(\mathrm{T}(n)\, g) = \langle g, \mathrm{T}(n)\rangle = \varphi(\mathrm{T}(n)) = a_n(f_\varphi)$$

for all $n \in \mathbb{N}$. □

Proposition 4.3 allows us to compute the q–expansions of the elements of a basis of $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ as long as we know a basis of $\mathrm{Hom}_{\mathbb{C}}(\mathbb{T}_{\mathbb{C}}, \mathbb{C})$. However, it is impossible to describe explicitly the action of Hecke operators on $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ without a precise description of $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ (and this is our ultimate goal). Thus, we are going to introduce another set of objects on which Hecke operators act in the same way. That is to say, we are going to compute the action of $\mathbb{T}$ using a space which contains an isomorphic copy of $\mathrm{S}_k(\Gamma_0(\mathrm{N}))$ and on which this action can be explicitly described in a simpler way.

To this aim, we restrict ourselves to the case in which $k = 2$. Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. By proposition 2.27 and lemma 2.29, there is an isomorphism (of complex vector spaces) between $\mathrm{S}_2(\Gamma)$ and the space of holomorphic differential 1–forms on $\mathrm{X}(\Gamma)$, which we denote by $\Omega^1(\mathrm{X}(\Gamma))$. Since $\mathrm{X}(\Gamma)$ is a compact Riemann surface, the complex dimension of $\Omega^1(\mathrm{X}(\Gamma))$ coincides with g, the genus of $\mathrm{X}(\Gamma)$. But the genus is a topological invariant: topologically, $\mathrm{X}(\Gamma)$ is a g–holed torus. Therefore, its first homology group $\mathrm{H}_1(\mathrm{X}(\Gamma), \mathbb{Z})$ is a free abelian group of rank $2g$ (there are two generators for each hole). We can now consider the real homology $\mathrm{H}_1(\mathrm{X}(\Gamma), \mathbb{R}) = \mathrm{H}_1(\mathrm{X}(\Gamma), \mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{R}$, which is a real vector space of dimension $2g$. Consequently,

$$\dim_{\mathbb{R}}(\mathrm{H}_1(\mathrm{X}(\Gamma), \mathbb{R})) = 2g = 2\dim_{\mathbb{C}}(\Omega^1(\mathrm{X}(\Gamma))) = \dim_{\mathbb{R}}(\Omega^1(\mathrm{X}(\Gamma)))\,.$$

In fact, the general theory of compact Riemann surfaces provides an explicit relation between $\mathrm{H}_1(\mathrm{X}(\Gamma), \mathbb{R})$ and $\Omega^1(\mathrm{X}(\Gamma))$.

Theorem 4.4. *The integration pairing*

$$\begin{aligned} \mathrm{I}\colon \mathrm{H}_1(\mathrm{X}(\Gamma),\mathbb{R})\times\Omega^1(\mathrm{X}(\Gamma)) &\longrightarrow \mathbb{C} \\ \Big(\sum_{i=1}^{r}\lambda_i[\gamma_i],\omega\Big) &\longmapsto \sum_{i=1}^{r}\lambda_i\int_{\gamma_i}\omega \end{aligned}$$

is non-degenerate. Consequently, it induces an isomorphism between $\mathrm{H}_1(\mathrm{X}(\Gamma),\mathbb{R})$ *and* $\mathrm{Hom}_{\mathbb{C}}(\Omega^1(\mathrm{X}(\Gamma)),\mathbb{C})$ *(both regarded as real vector spaces).*

Idea of the proof. First, observe that the pairing is well-defined because integrals over homologous paths coincide. (This result is a version of the Cauchy integral theorem for compact Riemann surfaces.) Indeed, if $\omega\in\Omega^1(\mathrm{X}(\Gamma))$, then $d\omega=0$ and so, by Stokes's theorem,

$$\int_{\partial\Delta}\omega=\int_{\Delta}d\omega=0$$

for every 1–boundary $\partial\Delta$. We always consider representatives of the homology classes which are (at least) rectifiable curves so that the integrals make sense.

Let $a_1,\dots,a_{2g}$ be the fundamental cycles relative to a polygonal decomposition of $\mathrm{X}(\Gamma)$. The elements of $\mathrm{H}_1(\mathrm{X}(\Gamma),\mathbb{R})$ are formal linear combinations

$$\sigma=\sum_{i=1}^{2g}\lambda_i[a_i]$$

with $\lambda_i\in\mathbb{R}$ for all i. Thus, the integration pairing is given by

$$\mathrm{I}(\sigma,\omega)=\sum_{i=1}^{2g}\lambda_i\int_{a_i}\omega\,.$$

Using the Abel–Jacobi theorem for compact Riemann surfaces, one can check that the $\mathbb{Z}$–span of the elements $\mathrm{I}([a_i],\cdot\,)$ (for $1\le i\le 2g$) is a lattice of maximal rank $2g$ in $\mathrm{Hom}_{\mathbb{C}}(\Omega^1(\mathrm{X}(\Gamma)),\mathbb{C})\cong\mathbb{C}^g\cong\mathbb{R}^{2g}$ (the quotient of $\mathrm{Hom}_{\mathbb{C}}(\Omega^1(\mathrm{X}(\Gamma)),\mathbb{C})$ by this lattice is precisely the Jacobian of $\mathrm{X}(\Gamma)$). In conclusion, the integration pairing I induces an isomorphism (of real vector spaces) between $\mathrm{H}_1(\mathrm{X}(\Gamma),\mathbb{R})$ and $\mathrm{Hom}_{\mathbb{C}}(\Omega^1(\mathrm{X}(\Gamma)),\mathbb{C})$. □

In this sense, the first homology group of $\mathrm{X}(\Gamma)$ with real coefficients is dual to $\mathrm{S}_2(\Gamma)$. That is why we are going to study the first homology group of $\mathrm{X}(\Gamma)$ (by

means of what are known as modular symbols) and how the action of Hecke operators translates to it. We explain the results involving Hecke operators only for $\Gamma_0(N)$ because we have an explicit description of $T(n)$ in this case.

Definition 4.5. Let $n \in \mathbb{N}$. The n–th *Hecke operator* acts on the first homology group of $X_0(N)$ with real coefficients, $T(n)\colon H_1(X_0(N),\mathbb{R}) \to H_1(X_0(N),\mathbb{R})$, in the following way: for every $\sigma \in H_1(X_0(N),\mathbb{R})$, $T(n)\sigma$ is the only element of $H_1(X_0(N),\mathbb{R})$ with the property that

$$I(T(n)\sigma,\omega) = I(\sigma,T(n)\omega)$$

for all $\omega \in \Omega^1(X_0(N))$, where I is the integration pairing described in theorem 4.4. (Here, we identify $\Omega^1(X_0(N))$ with $S_2(\Gamma_0(N))$.)

4.2 Homology and modular symbols

Throughout the remainder of this chapter, let Γ be a fixed congruence subgroup of $SL_2(\mathbb{Z})$ and let $\pi\colon \mathbb{H}^* \to X(\Gamma)$ be the natural projection map. We would like to obtain an explicit description of a simple set of generators of $H_1(X(\Gamma),\mathbb{Z})$. Such generators are (homology classes of) paths on $X(\Gamma)$. Since $X(\Gamma)$ is defined as a quotient of $\mathbb{H}^*$, we start by considering paths on $\mathbb{H}^*$ (whose images under π are paths on $X(\Gamma)$). Since we are going to integrate differential forms along these paths, we impose some additional conditions. That is, we are always going to choose "regular enough" representatives of the homology classes.

Let $r,s \in \mathbb{H}^*$. By a path joining r and s in $\mathbb{H}^*$, we mean a piecewise smooth (i.e., piecewise continuously differentiable) path lying inside $\mathbb{H}$ except for possibly the endpoints. Moreover, we require that the path be smooth at the endpoints in the following sense. If $s = \infty$, the path leading to ∞ should be contained in a vertical strip of finite width and its image under the map $z \mapsto e^{2\pi i z}$ should be a piecewise smooth path leading to 0. If s is any other cusp, there is an element of $SL_2(\mathbb{Z})$ which maps a neighbourhood of s to a neighbourhood of ∞ and we can define the condition of smoothness similarly. With this notion of path in $\mathbb{H}^*$, its projection on $X(\Gamma)$ is also a piecewise smooth path.

Lemma 4.6. *Let $r,s \in \mathbb{H}^*$. Any two paths on $\mathbb{H}^*$ joining r with s are homotopic. Consequently, the images of two such paths under π are also homotopic.*

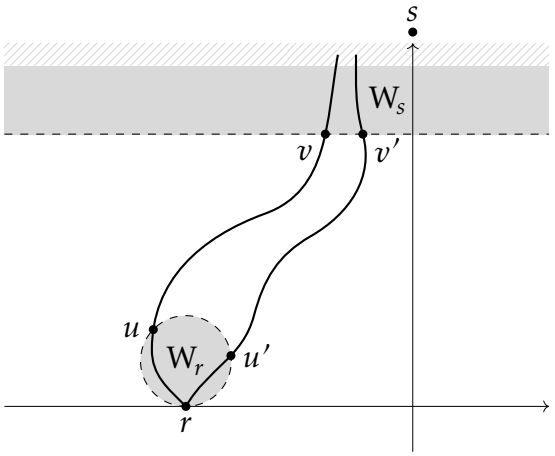


Figure 4.1: Two homotopic paths joining r and s.

Proof. Since $\mathbb{H}$ is simply connected, we only need to study the case in which at least one of r and s is a cusp. For instance, assume that $r \in \mathbb{Q}$ and $s = \infty$, as in figure 4.1. The smoothness condition at the endpoints of the two paths ensures the existence of fundamental neighbourhoods W_r of r and W_s of s which divide each path into three portions, as illustrated in figure 4.1: there is one portion lying in $\mathbb{H}$, away from the endpoints, and the two tails leading to r and to s. Since W_r is simply connected, the pieces between r and u and between r and u' are homotopic in W_r. Similarly, the pieces between v and s and between v' and s are homotopic in W_s because W_s is simply connected. And we have already observed that $\mathbb{H}$ is also simply connected, so the remaining pieces are homotopic as well. (In fact, all these open sets are not only simply connected, but also convex.)

Finally, since $\pi\colon \mathbb{H}^* \to X(\Gamma)$ is continuous, the composition of π with a homotopy in $\mathbb{H}^*$ is a homotopy in $X(\Gamma)$. □

Let $r, s \in \mathbb{H}^*$ and consider a path $P_{r,s}$ from r to s in $\mathbb{H}^*$. Lemma 4.6 implies that, for every $\omega \in \Omega^1(X(\Gamma))$, the integral

$$\int_r^s \pi^*(\omega) = \int_{P_{r,s}} \pi^*(\omega)$$

does not depend on the choice of the path $P_{r,s}$ (but only on its endpoints).

Definition 4.7. Let $r, s \in \mathbb{H}^*$. The *modular symbol* $\{r, s\}$ for Γ is the element of $H_1(X(\Gamma), \mathbb{R})$ corresponding to

$$\left(\omega \longmapsto \int_r^s \pi^*(\omega)\right) \in \mathrm{Hom}_{\mathbb{C}}(\Omega^1(X(\Gamma)), \mathbb{C})$$

under the integration pairing described in theorem 4.4.

Observe that the definition of modular symbol depends on the congruence subgroup Γ. For example, we could write $\{r,s\}_\Gamma$ instead of just $\{r,s\}$ to make it explicit. However, we are going to omit Γ from the notation because we are not going to mix modular symbols for different congruence subgroups.

In general, the modular symbol $\{r,s\}$ is an element of $H_1(X(\Gamma),\mathbb{R})$. But, if $\pi(r) = \pi(s)$, a path from r to s on $\mathbb{H}^*$ becomes a closed path on $X(\Gamma)$ and so $\{r,s\}$ can be regarded as an element of $H_1(X(\Gamma),\mathbb{Z})$ (viewed as a submodule of $H_1(X(\Gamma),\mathbb{R})$). The modular symbols which belong to the integral homology group are sometimes called integral modular symbols.

Actually, we will focus on the case in which r and s are cusps. Let $C(\Gamma)$ be the image of $\mathbb{P}^1_\mathbb{Q}$ in $X(\Gamma)$: we call its elements the cusps of $X(\Gamma)$, since they are the images of the cusps for Γ under π. (Also, we write $C_0(N) = C(\Gamma_0(N))$.) Thus, if $r,s \in \mathbb{P}^1_\mathbb{Q}$, $\{r,s\}$ can be viewed as an element of $H_1(X(\Gamma),C(\Gamma),\mathbb{Z})$, the first homology group of $X(\Gamma)$ relative to the cusps. Consequently, we think of $\{r,s\}$ as (the homology class of) a geodesic path from r to s in the Poincaré upper half-plane (regarded as a model of the hyperbolic plane), as illustrated in figure 4.2.

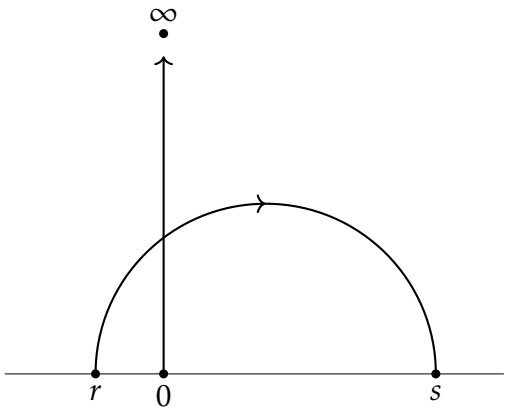


Figure 4.2: Geodesic paths from r to s and from 0 to ∞ in $\mathbb{H}^*$.

Definition 4.8. We define a left action of $GL_2^+(\mathbb{Q})$ on the space of modular symbols for Γ in the following way:

$$\alpha\{r,s\} = \{\alpha(r),\alpha(s)\}$$

for all $r,s \in \mathbb{H}^*$ and all $\alpha \in GL_2^+(\mathbb{Q})$ and this is extended by linearity.

Proposition 4.9. *Let $r, s, t \in \mathbb{H}^*$ and let $\gamma, \gamma' \in \Gamma$. We have the following identities:*

(1) $\{r, r\} = 0$;
(2) $\{r, s\} + \{s, r\} = 0$;
(3) $\{r, s\} + \{s, t\} + \{t, r\} = 0$;
(4) $\gamma\{r, s\} = \{r, s\}$;
(5) $\{r, \gamma(r)\} = \{s, \gamma(s)\}$;
(6) $\{r, (\gamma\gamma')(r)\} = \{r, \gamma(r)\} + \{r, \gamma'(r)\}$.

Proof. (1) and (2) are immediate from the definition of modular symbols and the properties of the integrals. For (3), observe that $\{r, s\} + \{s, t\} + \{t, r\}$ corresponds to integration of holomorphic functions along the boundary of a triangle with vertices r, s and t: these integrals are always 0 by the Cauchy integral theorem.

Let $P_{r,s}$ be a path from r to s in $\mathbb{H}^*$. Then, $\gamma(P_{r,s})$ is a path from $\gamma(r)$ to $\gamma(s)$ and the images of $P_{r,s}$ and $\gamma(P_{r,s})$ under π coincide because $\gamma \in \Gamma$. This completes the proof of (4).

Finally, (5) and (6) are a consequence of the previous identities. Indeed,

$$\{r, \gamma(r)\} = \{r, s\} + \{s, \gamma(s)\} + \{\gamma(s), \gamma(r)\} = \{r, s\} + \{s, \gamma(s)\} + \{s, r\} = \{s, \gamma(s)\}$$

and also

$$\{r, (\gamma\gamma')(r)\} = \{r, \gamma(r)\} + \{\gamma(r), \gamma(\gamma'(r))\} = \{r, \gamma(r)\} + \{r, \gamma'(r)\}. \qquad \square$$

Proposition 4.10. *Let $r \in \mathbb{H}^*$. The map*

$$\begin{aligned} \Phi : \overline{\Gamma} &\longrightarrow H_1(X(\Gamma), \mathbb{Z}) \\ \overline{\gamma} &\longmapsto \{r, \gamma(r)\} \end{aligned}$$

is a surjective group morphism which does not depend on the choice of r. Moreover, the kernel of Φ is generated by the images of the commutators, the elliptic elements and the parabolic elements of Γ in $\overline{\Gamma} = (\{\pm 1\} \cdot \Gamma)/\{\pm 1\}$.

Proof. The identities (5) and (6) of proposition 4.9 imply that Φ is a morphism which does not depend on the choice of r. To prove the remaining assertions, we must use a geometric interpretation of Φ.

We can assume that r is neither an elliptic point nor a cusp. Let $\mathbb{H}^0$ be the complement of the subset of elliptic points in $\mathbb{H}$ and let $Y^0(\Gamma) = \pi(\mathbb{H}^0)$. Since the ramification points of π are precisely the elliptic points and the cusps, $\mathbb{H}^0$ is a

(path-connected) covering space of $Y^0(\Gamma)$ with covering map $\pi|_{\mathbb{H}^0}\colon \mathbb{H}^0 \to Y^0(\Gamma)$. Therefore, $\pi|_{\mathbb{H}^0}$ induces a morphism Ψ from $\pi_1(Y^0(\Gamma), \pi(r))$ (the fundamental group of $Y^0(\Gamma)$ with base point $\pi(r)$) to $\overline{\Gamma}$ as follows. Every closed path P in $Y^0(\Gamma)$ starting at $\pi(r)$ has a unique lift to a path $\widetilde{P}$ in $\mathbb{H}^0$ starting at r. The other endpoint of $\widetilde{P}$ must be of the form $\gamma(r)$ for some $\gamma \in \Gamma$ whose image $\overline{\gamma}$ in $\overline{\Gamma}$ is uniquely determined. In this situation, Ψ sends the homotopy class of P to $\overline{\gamma}$. This morphism is well-defined because a homotopy of paths starting at $\pi(r)$ in $Y^0(\Gamma)$ has a unique lift to a homotopy of paths starting at r in $\mathbb{H}^0$. Furthermore, Ψ is surjective because $\mathbb{H}^0$ is path-connected and so we can always find a path from r to $\gamma(r)$ in $\mathbb{H}^0$.

The composite map $F = \Phi \circ \Psi\colon \pi_1(Y^0(\Gamma), \pi(r)) \to H_1(X(\Gamma), \mathbb{Z})$ sends the homotopy class of a closed path P starting at $\pi(r)$ in $Y^0(\Gamma)$ to the first homology class of P in $X(\Gamma)$ (intuitively, Ψ lifts P to a path $\widetilde{P}$ in $\mathbb{H}^0$ and Φ projects $\widetilde{P}$ on $X(\Gamma)$). That is to say, F coincides with the canonical morphism from the fundamental group of $Y^0(\Gamma)$ to the first homology group of its compactification $X(\Gamma)$. Hence, F factors through $H_1(Y^0(\Gamma), \mathbb{Z})$.

Since $Y^0(\Gamma)$ is path-connected (it is a g–holed torus minus a finite set of points), the canonical map $\pi_1(Y^0(\Gamma), \pi(r)) \to H_1(Y^0(\Gamma), \mathbb{Z})$ is surjective and its kernel is the commutator subgroup of $\pi_1(Y^0(\Gamma), \pi(r))$. Likewise, the natural map $H_1(Y^0(\Gamma), \mathbb{Z}) \to H_1(X(\Gamma), \mathbb{Z})$ is surjective and its kernel is generated by the cycles round the images of elliptic points and cusps under π (these cycles contract in the compactification). In conclusion, F is surjective and its kernel is generated by the commutator subgroup of $\pi_1(Y^0(\Gamma), \pi(r))$ and by the homotopy classes of closed paths round the images of elliptic points and cusps under π. Thus, Φ is also surjective and its kernel is the image under Ψ of the kernel of F.

On the one hand, the image of the commutator subgroup of $\pi_1(Y^0(\Gamma), \pi(r))$ under Ψ is generated by the image of the commutator subgroup of Γ in $\overline{\Gamma}$. On the other hand, (homotopy classes of) cycles round elliptic points or cusps correspond to the elements of Γ which fix these elliptic points or cusps (that is, the elliptic and parabolic elements of Γ), as we saw when we defined the charts of $X(\Gamma)$ at these points. This completes the description of the kernel of Φ. □

We are going to use the previous result, corresponding to proposition 1.4 of Manin's paper [5], to find a simple way to represent the elements of $H_1(X(\Gamma), \mathbb{Z})$.

Definition 4.11. The *distinguished modular symbols* for Γ are those of the form $\{\alpha(0), \alpha(\infty)\}$ for some $\alpha \in \mathrm{SL}_2(\mathbb{Z})$. If $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then $\{\alpha(0), \alpha(\infty)\} = \left\{\frac{b}{d}, \frac{a}{c}\right\}$.

The following result, known in the literature as Manin's trick, is proposition 1.6 of Manin's paper [5].

Theorem 4.12 (Manin). *Let $\overline{\alpha_1},\dots,\overline{\alpha_m}$ be a set of representatives of the right cosets of $\overline{\Gamma}$ in $\mathrm{PSL}_2(\mathbb{Z})$. Every element $[a]$ of $\mathrm{H}_1(\mathrm{X}(\Gamma),\mathbb{Z})$ can be represented as a $\mathbb{Z}$–linear combination of distinguished modular symbols of the form*

$$[a] = \sum_{j=1}^{m} \lambda_j \{\alpha_j(0), \alpha_j(\infty)\}$$

with the property that

$$\sum_{j=1}^{m} \lambda_j [\pi(\alpha_j(\infty)) - \pi(\alpha_j(0))] = 0$$

(as a 0–cycle on $\mathrm{X}(\Gamma)$).

Proof. First, observe that $\alpha_j\{0,\infty\}$ is independent of the representative of the right coset $\overline{\Gamma}\overline{\alpha_j}$, by the assertion (4) of proposition 4.9.

By proposition 4.10, we can write $[a] = \{0, \gamma(0)\}$ for some $\gamma \in \Gamma$. If $\gamma(0) = \infty$, this modular symbol is distinguished and $\pi(\infty) - \pi(0) = 0$. Otherwise, $\gamma(0) \in \mathbb{Q}$ and we can write $\gamma(0) = \frac{p}{q}$ in lowest terms, with $q > 0$. We expand

$$\frac{p}{q} = x_0 + \cfrac{1}{x_1 + \cfrac{1}{x_2 + \cfrac{1}{\ddots + \cfrac{1}{x_n}}}}$$

(as a finite continued fraction) and consider the successive convergents

$$\frac{p_{-2}}{q_{-2}} = \frac{0}{1}, \quad \frac{p_{-1}}{q_{-1}} = \frac{1}{0}, \quad \frac{p_0}{q_0} = \frac{p_0}{1}, \quad \dots, \quad \frac{p_n}{q_n} = \frac{p}{q}$$

(all of them written in lowest terms and the first two included formally). In this situation, we can express

$$\left\{0, \frac{p}{q}\right\} = \sum_{k=-1}^{n} \left\{\frac{p_{k-1}}{q_{k-1}}, \frac{p_k}{q_k}\right\},$$

by proposition 4.9. This representation satisfies the required property because $\pi(0) = \pi(\gamma(0))$. Thus, we only need to prove that the modular symbols $\left\{\frac{p_{k-1}}{q_{k-1}}, \frac{p_k}{q_k}\right\}$ are distinguished. But it is well-known that $p_k q_{k-1} - p_{k-1} q_k = (-1)^{k-1}$ and so

$$\beta_k = \begin{pmatrix} p_k & (-1)^{k-1} p_{k-1} \\ q_k & (-1)^{k-1} q_{k-1} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

Consequently, we can express

$$\left\{\frac{p_{k-1}}{q_{k-1}}, \frac{p_k}{q_k}\right\} = \{\beta_k(0), \beta_k(\infty)\} = \{\alpha_{j_k}(0), \alpha_{j_k}(\infty)\}$$

for some $j_k \in \{1, \dots, m\}$. All these modular symbols (for $k \in \{-1, 0, \dots, n\}$) are distinguished. □

Proposition 4.13. *Every $\mathbb{Z}$–linear combination of modular symbols of the form*

$$c = \sum_{j=1}^{n} \lambda_j \{r_j, s_j\}$$

with the property that

$$\partial c = \sum_{j=1}^{n} \lambda_j [\pi(s_j) - \pi(r_j)] = 0$$

(as a 0–cycle on $\mathrm{X}(\Gamma)$) is an element of the first integral homology group $\mathrm{H}_1(\mathrm{X}(\Gamma), \mathbb{Z})$.

Proof. We have to prove that c is a $\mathbb{Z}$–linear combination of homology classes of closed paths (i.e., of 1–cycles).

We may assume that $\lambda_j = \pm 1$ for all $j \in \{1, \dots, n\}$ by allowing repetitions. In fact, we may assume that $\lambda_j = 1$ for all $j \in \{1, \dots, n\}$ because $\{s, r\} = -\{r, s\}$. Now, we reorder the sum and express

$$c = \sum_{i=1}^{u} c_i = \sum_{i=1}^{u} \sum_{j=n_{i-1}+1}^{n_i} \{r_j, s_j\},$$

where $0 = n_0 < n_1 < \cdots < n_u = n$, so that $\pi(s_{j-1}) = \pi(r_j)$ for all $j \in \{n_{i-1} + 2, \dots, n_i\}$ and $\partial c_i = 0$ for all $i \in \{1, \dots, u\}$. We can do so inductively: at the l–th step, if

$$\partial\Big(\sum_{j=n_{i-1}+1}^{l-1} \{r_j, s_j\} \Big) = 0,$$

we set $n_i = l - 1$ and choose $\{r_l, s_l\}$ to be any of the remaining modular symbols; otherwise, the coefficient of $\pi(s_{l-1})$ in the preceding sum must be $+1$ and so we can choose $\{r_l, s_l\}$, among the modular symbols which have not been selected in the previous steps, satisfying that $\pi(r_l) = \pi(s_{l-1})$.

In conclusion, for each $i \in \{1, \dots, u\}$, c_i is the homology class of a closed path in $X(\Gamma)$ obtained by concatenating the images under π of the geodesic paths (in $\mathbb{H}^*$) from r_j to s_j for $j \in \{n_{i-1} + 1, \dots, n_i\}$. Therefore, $c \in H_1(X(\Gamma), \mathbb{Z})$. □

The previous results provide a way to represent the first homology classes of $X(\Gamma)$ in terms of a set of representatives for $\overline{\Gamma} \backslash \mathrm{PSL}_2(\mathbb{Z})$. We are going to simplify this representation even more by means of what are known as Manin symbols.

4.3 Manin symbols

Consider the matrices

$$\sigma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \tau = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}.$$

Observe that, with the notation of theorem 1.9, $\sigma = S$ and $\tau = TS$. In particular, these two matrices generate $\mathrm{SL}_2(\mathbb{Z})$. Moreover, $\sigma^2 = \tau^3 = -1$.

Since $\sigma\{0, \infty\} = \{\infty, 0\}$ and $\tau\{0, \infty\} = \{\infty, 1\}$ and $\tau\{\infty, 1\} = \{1, 0\}$,

$$\alpha\{0, \infty\} + \alpha\sigma\{0, \infty\} = 0 \quad \text{and} \quad \alpha\{0, \infty\} + \alpha\tau\{0, \infty\} + \alpha\tau^2\{0, \infty\} = 0$$

for every $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ (by proposition 4.9). We are going to see that this system of relations is complete in some sense.

Definition 4.14. Let $\overline{\alpha_1}, \dots, \overline{\alpha_m}$ be a set of representatives of the right cosets of $\overline{\Gamma}$ in $\mathrm{PSL}_2(\mathbb{Z})$. The formal symbols $(\overline{\alpha_j})$ for $j \in \{1, \dots, m\}$ are called *Manin symbols* for Γ. We extend this notation and write $(\overline{\alpha}) = (\overline{\Gamma}\,\overline{\alpha_j}) = (\overline{\alpha_j})$ for all $\overline{\alpha} \in \overline{\Gamma}\,\overline{\alpha_j}$. (There is one Manin symbol for each right coset, but we identify these right cosets with their representatives.)

Definition 4.15. We define a right action of $\mathrm{SL}_2(\mathbb{Z})$ on the set of Manin symbols for Γ in the following way: for all $\overline{\alpha} \in \mathrm{PSL}_2(\mathbb{Z})$ and all $\beta \in \mathrm{SL}_2(\mathbb{Z})$, $(\overline{\alpha})\beta = (\overline{\alpha\beta})$.

Definition 4.16. Let $\mathrm{Man}(\Gamma)$ be the free abelian group generated by the Manin symbols $(\overline{\alpha})$ for Γ. The group of 1–*chains of Manin symbols* for Γ is the quotient

$C(\mathrm{Man}(\Gamma))$ of $\mathrm{Man}(\Gamma)$ by the subgroup generated by the elements of the form $(\overline{\alpha}) + (\overline{\alpha})\sigma$ and by the elements $(\overline{\alpha})$ such that $(\overline{\alpha}) = (\overline{\alpha})\sigma$. The *boundary* of a Manin symbol $(\overline{\alpha})$ is the element $\pi(\alpha(\infty)) - \pi(\alpha(0))$ of the free abelian group generated by the cusps $C(\Gamma)$; we extend this boundary operator by $\mathbb{Z}$–linearity to the entire group $C(\mathrm{Man}(\Gamma))$. Its kernel $Z(\mathrm{Man}(\Gamma))$ is called the group of 1–*cycles of Manin symbols* for Γ. The group of 1–*boundaries of Manin symbols* for Γ is the subgroup $B(\mathrm{Man}(\Gamma))$ of $C(\mathrm{Man}(\Gamma))$ generated by the elements $(\overline{\alpha}) + (\overline{\alpha})\tau + (\overline{\alpha})\tau^2$ and by the elements $(\overline{\alpha})$ such that $(\overline{\alpha}) = (\overline{\alpha})\tau$.

In the previous definition, the boundary of a 1–chain of Manin symbols for Γ is well-defined because σ interchanges 0 and ∞ and, in $C(\mathrm{Man}(\Gamma))$, $(\overline{\alpha})\sigma = -(\overline{\alpha})$. (We identify the elements of $C(\mathrm{Man}(\Gamma))$ with their representatives.)

Lemma 4.17. $B(\mathrm{Man}(\Gamma))$ *is a subgroup of* $Z(\mathrm{Man}(\Gamma))$.

Proof. Let $(\overline{\alpha})$ be a Manin symbol. If $(\overline{\alpha}) = (\overline{\alpha})\tau$, then $\Gamma\alpha(0) = \Gamma\alpha(\tau(0)) = \Gamma\alpha(\infty)$ and so the boundary of $(\overline{\alpha})$ is $\pi(\alpha(\infty)) - \pi(\alpha(0)) = 0$. Therefore, $(\overline{\alpha}) \in Z(\mathrm{Man}(\Gamma))$. Also, the element $(\overline{\alpha}) + (\overline{\alpha})\tau + (\overline{\alpha})\tau^2$ belongs to $Z(\mathrm{Man}(\Gamma))$ because its boundary is $\pi(\alpha(\infty)) - \pi(\alpha(0)) + \pi(\alpha(1)) - \pi(\alpha(\infty)) + \pi(\alpha(0)) - \pi(\alpha(1)) = 0$. □

Lemma 4.18. *The morphism*

$$\begin{aligned} \xi\colon Z(\mathrm{Man}(\Gamma))/B(\mathrm{Man}(\Gamma)) &\longrightarrow H_1(X(\Gamma),\mathbb{Z}) \\ \sum_{j=1}^{m} \lambda_j(\overline{\alpha_j}) + B(\mathrm{Man}(\Gamma)) &\longmapsto \sum_{j=1}^{m} \lambda_j\{\alpha_j(0), \alpha_j(\infty)\} \end{aligned}$$

is well-defined and surjective.

Proof. First, observe that proposition 4.13 implies that the image of ξ is contained in $H_1(X(\Gamma),\mathbb{Z})$ (as long as ξ does not depend on the choice of the representative of each element of the quotient $Z(\mathrm{Man}(\Gamma))/B(\mathrm{Man}(\Gamma))$). Second,

$$\alpha\{0,\infty\} + \alpha\tau\{0,\infty\} + \alpha\tau^2\{0,\infty\} = \{\alpha(0),\alpha(\infty)\} + \{\alpha(\infty),\alpha(1)\} + \{\alpha(1),\alpha(0)\} = 0\,.$$

Finally, if $(\overline{\alpha}) = (\overline{\alpha})\tau$, then $3\alpha\{0,\infty\} = \alpha\{0,\infty\} + \alpha\tau\{0,\infty\} + \alpha\tau^2\{0,\infty\} = 0$ and, since $H_1(X(\Gamma),\mathbb{R})$ is torsion-free, $\{\alpha(0),\alpha(\infty)\} = 0$. In conclusion, ξ is well-defined. The surjectivity of ξ is a direct consequence of theorem 4.12. □

The following theorem is the main result of this chapter, as it gives an algebraic presentation of the group $H_1(X(\Gamma),\mathbb{Z})$ which is very convenient for

computations. It corresponds to theorem 1.9 of Manin's paper [5] (the proof here is essentially the same, but some steps are explained in more detail).

Theorem 4.19 (Manin). *The map*

$$\begin{aligned} \xi\colon Z(\mathrm{Man}(\Gamma))/B(\mathrm{Man}(\Gamma)) &\longrightarrow H_1(X(\Gamma),\mathbb{Z}) \\ \sum_{j=1}^{m} \lambda_j(\overline{\alpha_j}) + B(\mathrm{Man}(\Gamma)) &\longmapsto \sum_{j=1}^{m} \lambda_j\{\alpha_j(0),\alpha_j(\infty)\} \end{aligned}$$

is an isomorphism.

Proof. Lemma 4.18 asserts that ξ is surjective, so we have to prove that it is also injective. To this aim, we are going to triangulate $X(\Gamma)$ in order to obtain a cell complex L with homology $Z_1(L)/B_1(L)$ (the quotient of the 1–cycles of L by the 1–boundaries of L) coinciding with $H_1(X(\Gamma),\mathbb{Z})$. Then, we are going to embed $Z(\mathrm{Man}(\Gamma))/B(\mathrm{Man}(\Gamma))$ in $Z_1(L)/B_1(L)$.

For any two points $r,s\in\mathbb{H}^*$, let $\langle r,s\rangle$ denote the segment joining r and s along the geodesic in $\mathbb{H}$ oriented from r to s (regarding $\mathbb{H}$ as a model of the hyperbolic plane: geodesics are thus semicircles and lines orthogonal to the real axis). The polygons appearing in this proof are going to be formed by geodesic segments in $\mathbb{H}^*$ joining the vertices of these figures; we also consider their images under the projection π.

Let E be the interior of the triangle with vertices $\{0,1,i\infty\}$, shown in figure 4.3. Let E′ be the union of the interior of the quadrilateral with vertices $\{i,\rho,1+i,i\infty\}$ and the side $\langle i,\rho\rangle$ except for the vertex i. By theorem 1.9, the closure of E′ is a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$. Moreover, none of the sides of this quadrilateral contains two distinct points which are identified under the action of $\mathrm{SL}_2(\mathbb{Z})$ (so neither under the action of Γ). Therefore, each of these sides can be embedded in $X(\Gamma)$.

In E, we have the three regions E′, τ(E′) and τ^2(E′) as illustrated in figure 4.3. The closure of each of them is a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$. For instance, the closure of E′ can be obtained from $F=\left\{z\in\mathbb{H}:|z|\geq 1 \text{ and } |\mathfrak{Re}(z)|\leq\tfrac{1}{2}\right\}$ (which is a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$, by theorem 1.9) by translating the left half of F under $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$. In addition, each of the 1–simplices appearing in the boundaries of E′, τ(E′) and τ^2(E′) (which are the half-sides and half-medians of the triangle E) can be embedded in $X(\Gamma)$ because there are no self-identifications. In fact, no two distinct points of one of these half-sides and half-medians are $\mathrm{SL}_2(\mathbb{Z})$–equivalent, as can be checked using theorem 1.9. Indeed, the half-side

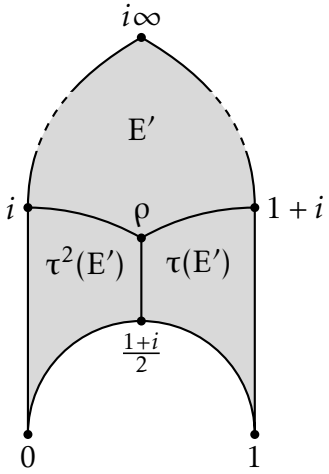


Figure 4.3: Triangulation of the region E.

$\langle i\infty, i\rangle$ is contained in $\mathring{F}$ (except for its endpoints) and the half-median $\langle \rho, i\rangle$ is on an edge of F with no self-identifications. The other half-sides and half-medians are $\mathrm{SL}_2(\mathbb{Z})$–equivalent to these: on the one hand, the half-sides $\langle 0, i\rangle$, $\langle 1, 1+i\rangle$, $\langle i\infty, 1+i\rangle$, $\langle 0, \frac{1+i}{2}\rangle$ and $\langle 1, \frac{1+i}{2}\rangle$ are the images of $\langle i\infty, i\rangle$ under the linear fractional transformations associated with $\sigma = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $\tau = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $\tau\sigma = \left(\begin{smallmatrix} -1 & -1 \\ 0 & -1 \end{smallmatrix}\right)$, $\tau^2 = \left(\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right)$ and $\tau^2\sigma = \left(\begin{smallmatrix} -1 & 0 \\ -1 & -1 \end{smallmatrix}\right)$, respectively; on the other hand, the half-medians $\langle \rho, 1+i\rangle$ and $\langle \rho, \frac{1+i}{2}\rangle$ are the images of $\langle \rho, i\rangle$ under the linear fractional transformations associated with $\tau = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $\tau^2 = \left(\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right)$, respectively.

We are in a position to describe the cell complex L.

The 0–cells of L are the images in $X(\Gamma)$ of the cusps and the points which are $\mathrm{SL}_2(\mathbb{Z})$–equivalent to i; that is, the points of $\pi(\mathbb{P}^1_{\mathbb{Q}}) \cup \pi(\mathrm{SL}_2(\mathbb{Z})i)$. Observe that these are the images under π of the vertices and the midpoints of the sides of the triangles $\alpha(E)$ for $\alpha \in \mathrm{SL}_2(\mathbb{Z})$.

The 1–cells of L are the images under π of the half-sides of the triangles $\alpha(E)$ for $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ oriented from the vertices to the midpoints (i.e., from the cusps to the points of $\mathrm{SL}_2(\mathbb{Z})i$). For every right coset $\overline{\Gamma}\,\overline{\alpha}$ in $\mathrm{PSL}_2(\mathbb{Z})$, we define $e_1(\overline{\Gamma}\,\overline{\alpha})$ to be the image of the path $\langle \alpha(\infty), \alpha(i)\rangle$ under π. Since all the half-sides of E are of the form $\langle \alpha(\infty), \alpha(i)\rangle$ for some $\alpha \in \mathrm{SL}_2(\mathbb{Z})$, every 1–cell can be expressed as $e_1(\overline{\Gamma}\,\overline{\alpha})$ for some right coset $\overline{\Gamma}\,\overline{\alpha}$. By definition,

$$\partial e_1(\overline{\Gamma}\,\overline{\alpha}) = \pi(\alpha(i)) - \pi(\alpha(\infty)).$$

Moreover, if $\overline{\Gamma}\,\overline{\alpha}$ and $\overline{\Gamma}\,\overline{\beta}$ are two right cosets such that $e_1(\overline{\Gamma}\,\overline{\alpha}) = e_1(\overline{\Gamma}\,\overline{\beta})$, then $\overline{\Gamma}\,\overline{\beta} = \overline{\Gamma}\,\overline{\alpha}$. Indeed, since $\mathrm{PSL}_2(\mathbb{Z})_i = \{\overline{1}, \overline{\sigma}\}$ and $\pi(\beta(i)) = \pi(\alpha(i))$, either $\overline{\Gamma}\,\overline{\beta} = \overline{\Gamma}\,\overline{\alpha}$ or

$\overline{\Gamma}\,\overline{\beta} = \overline{\Gamma}\,\overline{\alpha\sigma}$ (otherwise, the endpoints of $e_1(\overline{\Gamma}\,\overline{\beta})$ and of $e_1(\overline{\Gamma}\,\overline{\alpha})$ would be different). But, if $\overline{\Gamma}\,\overline{\alpha} \neq \overline{\Gamma}\,\overline{\alpha\sigma}$, the images of $\alpha(\mathring{\mathrm{F}})$ and of $\alpha\sigma(\mathring{\mathrm{F}})$ under π are disjoint (see the proof of proposition 1.10) and so the images of the paths $\langle\alpha(\infty), \alpha(i)\rangle$ and $\langle\alpha\sigma(\infty), \alpha\sigma(i)\rangle$ cannot coincide.

There are two types of 2–cells in L: those with 2 sides and those with 3. They are defined as follows.

For every right coset $\overline{\Gamma}\,\overline{\alpha}$ such that $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\tau}$, we have a 2–sided 2–cell $e_2(\overline{\Gamma}\,\overline{\alpha})$ which is $\pi(\alpha(\mathrm{E}'))$ with the usual orientation (induced by the positive orientation of the complex plane). Since $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\tau}$, the image of the half-median $\langle\alpha(\rho), \alpha(i)\rangle$ is a line from the centre to the boundary of $e_2(\overline{\Gamma}\,\overline{\alpha})$. Thus,

$$\begin{aligned}
\partial e_2(\overline{\Gamma}\,\overline{\alpha}) &= \pi(\langle\alpha(\infty), \alpha(i)\rangle + \langle\alpha(i), \alpha(\rho)\rangle + \langle\alpha(\rho), \alpha(1+i)\rangle + \langle\alpha(1+i), \alpha(\infty)\rangle) \\
&= \pi(\langle\alpha(\infty), \alpha(i)\rangle) - \pi(\langle\alpha(\infty), \alpha(1+i)\rangle) = e_1(\overline{\Gamma}\,\overline{\alpha}) - \pi(\langle\alpha\tau^2(\infty), \alpha\tau^2(1+i)\rangle) \\
&= e_1(\overline{\Gamma}\,\overline{\alpha}) - \pi(\langle\alpha(0), \alpha(i)\rangle) = e_1(\overline{\Gamma}\,\overline{\alpha}) - \pi(\langle\alpha\sigma(\infty), \alpha\sigma(i)\rangle) \\
&= e_1(\overline{\Gamma}\,\overline{\alpha}) - e_1(\overline{\Gamma}\,\overline{\alpha\sigma}).
\end{aligned}$$

If $\overline{\Gamma}\,\overline{\beta}$ is another right coset with the property that $\overline{\Gamma}\,\overline{\beta\tau} = \overline{\Gamma}\,\overline{\beta}$ and $e_2(\overline{\Gamma}\,\overline{\alpha}) = e_2(\overline{\Gamma}\,\overline{\beta})$, then $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\beta}$ because E′ is a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$ (see the proof of proposition 1.10).

For every right coset $\overline{\Gamma}\,\overline{\alpha}$ such that $\overline{\Gamma}\,\overline{\alpha} \neq \overline{\Gamma}\,\overline{\alpha\tau}$, we have a 3–sided 2–cell $e_2(\overline{\Gamma}\,\overline{\alpha})$ which is $\pi(\alpha(\mathrm{E}))$ with the usual orientation (induced by the positive orientation of the complex plane). In this case, the images of the three triangles $\alpha(\mathrm{E}')$, $\alpha\tau(\mathrm{E}')$ and $\alpha\tau^2(\mathrm{E}')$ are distinct and $e_2(\overline{\Gamma}\,\overline{\alpha})$ is their union. One checks easily that

$$\partial e_2(\overline{\Gamma}\,\overline{\alpha}) = \sum_{j=0}^{2} [e_1(\overline{\Gamma}\,\overline{\alpha\tau^j}) - e_1(\overline{\Gamma}\,\overline{\alpha\tau^j\sigma})].$$

Since E′ is a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$, $e_2(\overline{\Gamma}\,\overline{\beta}) = e_2(\overline{\Gamma}\,\overline{\alpha})$ is only possible for $\overline{\Gamma}\,\overline{\beta} = \overline{\Gamma}\,\overline{\alpha\tau^j}$ with $j \in \{0, 1, 2\}$: this is the only way to make $\beta(\mathrm{E}')$ Γ–equivalent to $\alpha(\mathrm{E}')$, $\alpha\tau(\mathrm{E}')$ or $\alpha\tau^2(\mathrm{E}')$ (see the proof of proposition 1.10).

We have thus defined a cell complex L which is a triangulation of the surface $\mathrm{X}(\Gamma)$. Indeed, by proposition 1.10, the translates of $\overline{\mathrm{E}'}$ by the right cosets of $\overline{\Gamma}$ in $\mathrm{PSL}_2(\mathbb{Z})$ form a fundamental domain for Γ. Consequently, $\mathrm{H}_1(\mathrm{X}(\Gamma), \mathbb{Z})$ is isomorphic to the quotient of the group $\mathrm{Z}_1(\mathrm{L})$ of 1–cycles of L by the group $\mathrm{B}_1(\mathrm{L})$ of 1–boundaries of L.

Let $\mathrm{C}_1(\mathrm{L})$ be the group of 1–chains of the complex L. We are going to prove

that the morphism

$$\begin{aligned}\varphi\colon \mathrm{C}(\mathrm{Man}(\Gamma)) &\longrightarrow \mathrm{C}_1(\mathrm{L})\\ (\overline{\alpha}) &\longmapsto e_1(\overline{\Gamma}\,\overline{\alpha\sigma}) - e_1(\overline{\Gamma}\,\overline{\alpha})\end{aligned}$$

induces an injective morphism from $\mathrm{Z}(\mathrm{Man}(\Gamma))/\mathrm{B}(\mathrm{Man}(\Gamma))$ to $\mathrm{Z}_1(\mathrm{L})/\mathrm{B}_1(\mathrm{L})$ which coincides with ξ.

First, observe that φ is well-defined. Indeed, for every Manin symbol $(\overline{\alpha})$, $\varphi((\overline{\alpha}) + (\overline{\alpha})\sigma) = 0$. Moreover, if $(\overline{\alpha}) = (\overline{\alpha})\sigma$, $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\sigma}$ and so $\varphi((\overline{\alpha})) = 0 = \varphi((\overline{\alpha\sigma}))$.

Second, we prove that φ is injective. Consider a 1–chain of Manin symbols

$$c = \sum_{\overline{\Gamma}\,\overline{\alpha}} n_{(\overline{\alpha})}(\overline{\alpha})$$

(where the sum is over the right cosets of $\overline{\Gamma}$ in $\mathrm{PSL}_2(\mathbb{Z})$) and assume further that this expression is normalised in the sense that $n_{(\overline{\alpha})}n_{(\overline{\alpha\sigma})} = 0$ for every Manin symbol $(\overline{\alpha})$: we can do so in view of the relations $(\overline{\alpha}) + (\overline{\alpha\sigma}) = 0$ and $(\overline{\alpha}) = 0$ if $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\sigma}$ in $\mathrm{C}(\mathrm{Man}(\Gamma))$. Then,

$$\varphi(c) = \sum_{\overline{\Gamma}\,\overline{\alpha}} n_{(\overline{\alpha})}[e_1(\overline{\Gamma}\,\overline{\alpha\sigma}) - e_1(\overline{\Gamma}\,\overline{\alpha})].$$

If $c \neq 0$, there is some $\overline{\Gamma}\,\overline{\alpha}$ such that $n_{(\overline{\alpha})} \neq 0$. In this case, $\overline{\Gamma}\,\overline{\alpha} \neq \overline{\Gamma}\,\overline{\alpha\sigma}$, which implies that $e_1(\overline{\Gamma}\,\overline{\alpha}) \neq e_1(\overline{\Gamma}\,\overline{\alpha\sigma})$. Similarly, if $\overline{\Gamma}\,\overline{\beta}$ is another right coset (distinct from $\overline{\Gamma}\,\overline{\alpha}$) with $n_{(\overline{\beta})} \neq 0$, the 1–cells $e_1(\overline{\Gamma}\,\overline{\alpha})$, $e_1(\overline{\Gamma}\,\overline{\alpha\sigma})$, $e_1(\overline{\Gamma}\,\overline{\beta})$ and $e_1(\overline{\Gamma}\,\overline{\beta\sigma})$ are all distinct. As a consequence, $\varphi(c) \neq 0$.

Third, φ preserves the boundaries. On the one hand, if $(\overline{\alpha})$ is a Manin symbol, $\partial\varphi((\overline{\alpha})) = \pi(\alpha\sigma(i)) - \pi(\alpha\sigma(\infty)) - \pi(\alpha(i)) + \pi(\alpha(\infty)) = \pi(\alpha(\infty)) - \pi(\alpha(0)) = \partial(\overline{\alpha})$. In particular, this means that $\varphi(\mathrm{Z}(\mathrm{Man}(\Gamma))) \subseteq \mathrm{Z}_1(\mathrm{L})$. On the other hand, $\mathrm{B}(\mathrm{Man}(\Gamma))$ is generated by the elements $(\overline{\alpha})$ for the right cosets $\overline{\Gamma}\,\overline{\alpha}$ such that $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\tau}$ and by the elements $(\overline{\alpha}) + (\overline{\alpha\tau}) + (\overline{\alpha\tau^2})$ for the right cosets $\overline{\Gamma}\,\overline{\alpha}$ such that $\overline{\Gamma}\,\overline{\alpha} \neq \overline{\Gamma}\,\overline{\alpha\tau}$ and, in turn, $\mathrm{B}_1(\mathrm{L})$ is generated by the elements $\partial e_2(\overline{\Gamma}\,\overline{\alpha}) = -\varphi((\overline{\alpha}))$ for the right cosets $\overline{\Gamma}\,\overline{\alpha}$ such that $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\tau}$ and by the elements $\partial e_2(\overline{\Gamma}\,\overline{\alpha}) = -\varphi((\overline{\alpha}) + (\overline{\alpha\tau}) + (\overline{\alpha\tau^2}))$ for the right cosets $\overline{\Gamma}\,\overline{\alpha}$ such that $\overline{\Gamma}\,\overline{\alpha} \neq \overline{\Gamma}\,\overline{\alpha\tau}$. Thus, $\mathrm{B}_1(\mathrm{L}) = \varphi(\mathrm{B}(\mathrm{Man}(\Gamma))) \subseteq \varphi(\mathrm{Z}(\mathrm{Man}(\Gamma)))$.

All in all, φ induces a monomorphism $\widetilde{\varphi}$ from $\mathrm{Z}(\mathrm{Man}(\Gamma))/\mathrm{B}(\mathrm{Man}(\Gamma))$ to the first homology group $\mathrm{Z}_1(\mathrm{L})/\mathrm{B}_1(\mathrm{L}) \cong \mathrm{H}_1(\mathrm{X}(\Gamma), \mathbb{Z})$. Finally, for every Manin symbol $(\overline{\alpha})$, the 1–chain $\varphi((\overline{\alpha}))$ is homologous to the path corresponding to $\{\alpha(0), \alpha(\infty)\}$. That is to say, the boundary of the image under π of the (degenerate) triangle with

vertices $\alpha(0)$, $\alpha(i)$ and $\alpha(\infty)$ is $\varphi((\overline{\alpha})) + \pi(\langle \alpha(\infty), \alpha(0) \rangle)$. From this, we deduce that $\varphi((\overline{\alpha})) + B_1(L) = \{\alpha(0), \alpha(\infty)\}$ in $C_1(L)/B_1(L)$. Extending this result by linearity, we obtain that $\widetilde{\varphi}$ coincides with ξ. □

The last result provides a purely algebraic description of the first homology group $H_1(X(\Gamma), \mathbb{Z})$ (in terms of generators and relations). If we focus on the case in which $\Gamma = \Gamma_0(N)$, we can give an even simpler presentation.

Let A be a commutative ring with identity. Recall that the projective line $\mathbb{P}^1_A$ is the quotient of the set $\{(a, b) \in A \times A : aA + bA = A\}$ by the equivalence relation $\sim$ defined by $(a, b) \sim (\lambda a, \lambda b)$ for all $\lambda \in A^\times$. Thus, the elements of $\mathbb{P}^1_A$ are of the form $(a : b)$ for $a, b \in A$ with the property that the ideal generated by a and b is the whole ring A (and with the convention that $(a : b) = (\lambda a : \lambda b)$ for all $\lambda \in A^\times$; these are projective coordinates).

Lemma 4.20. *For $j \in \{1, 2\}$, let $\alpha_j = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. The following conditions are equivalent:*

(a) $\Gamma_0(N)\alpha_1 = \Gamma_0(N)\alpha_2$;

(b) $c_1 d_2 \equiv c_2 d_1 \pmod{N}$;

(c) *there exists $\lambda \in \mathbb{Z}$ with $(\lambda, N) = 1$ such that $c_1 \equiv \lambda c_2$ and $d_1 \equiv \lambda d_2 \pmod{N}$.*

Proof. We compute the matrix

$$\alpha_1 \alpha_2^{-1} = \begin{pmatrix} a_1 d_2 - b_1 c_2 & b_1 a_2 - a_1 b_2 \\ c_1 d_2 - d_1 c_2 & d_1 a_2 - c_1 b_2 \end{pmatrix}.$$

Looking at its entries, we see that $\alpha_1 \alpha_2^{-1}$ is an element of $\Gamma_0(N)$ if and only if $c_1 d_2 - d_1 c_2 \equiv 0 \pmod{N}$, which means that (a) and (b) are equivalent.

Now we prove that (a) $\Longrightarrow$ (c). Since $\alpha_1 \alpha_2^{-1} \in \Gamma_0(N)$, the element $\lambda = d_1 a_2 - c_1 b_2$ which appears in the diagonal of $\alpha_1 \alpha_2^{-1}$ must be prime to N (because (λ, N) divides $\det(\alpha_1 \alpha_2^{-1}) = 1$). Using (b), we deduce that

$$\lambda c_2 = a_2 d_1 c_2 - b_2 c_1 c_2 \equiv a_2 d_2 c_1 - b_2 c_2 c_1 = c_1 \pmod{N}$$

and, similarly,

$$\lambda d_2 = a_2 d_1 d_2 - b_2 c_1 d_2 \equiv a_2 d_2 d_1 - b_2 c_2 d_1 = d_1 \pmod{N}.$$

Finally, it is obvious that (c) $\Longrightarrow$ (b). □

Proposition 4.21. *There is a bijection between the set of right cosets of* $\Gamma_0(N)$ *in* $\mathrm{SL}_2(\mathbb{Z})$ *and* $\mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$ *given by*

$$\Gamma_0(N)\begin{pmatrix} a & b \\ c & d \end{pmatrix} \longmapsto (c : d)$$

for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. *(Equivalently, since* $-1 \in \Gamma_0(N)$*, this map gives a bijection between the set of right cosets of* $\overline{\Gamma_0(N)}$ *in* $\mathrm{PSL}_2(\mathbb{Z})$ *and* $\mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$*.)*

Proof. Lemma 4.20 implies that the map described in the statement of this proposition is well-defined and injective, so we only need to prove that it is surjective.

If $(c : d) \in \mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$, the greatest common divisor of c, d and N is 1 (that is to say, the ideal generated by $c + N\mathbb{Z}$ and $d + N\mathbb{Z}$ is the whole $\mathbb{Z}/N\mathbb{Z}$). Moreover, the integers c and d are only determined modulo N. Let $M = (c, d)$ and write $c = Mc_0$ and $d = Md_0$. Since $(M, N) = 1$, there exist $x, y \in \mathbb{Z}$ such that $xM + yN = 1$ (Bézout's identity). In this situation,

$$(c : d) = (xc : xd) = (xc + yNc_0 : xd + yNd_0) = (c_0 : d_0)$$

and $(c_0, d_0) = 1$ by definition. Again, there exist $a, b \in \mathbb{Z}$ such that $ad_0 - bc_0 = 1$ and so $\alpha = \left(\begin{smallmatrix} a & b \\ c_0 & d_0 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. In conclusion, $\Gamma_0(N)\alpha$ is mapped to $(c : d)$. □

One can now translate all the previous results expressed in terms of Manin symbols to the set $\mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$ and make all the computations there.

Definition 4.22. There is a right action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$ given by

$$(c : d)\alpha = (cp + dr : cq + ds)$$

for all $(c : d) \in \mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$ and all $\alpha = \left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$.

The isomorphism described in theorem 4.19 can be expressed in terms of the symbols $(c : d)$ with the appropriate relations.

Chapter 5

Computations and examples

The previous chapter contains many allusions to the great importance of modular symbols for computations related to the theory of modular forms and Hecke operators. However, these computations have not been clearly exemplified, but only briefly mentioned. This chapter addresses this gap by means of particular examples and concrete computations.

In the first part of this chapter, we introduce an alternative presentation of the spaces of modular symbols, which is more appropriate for practical purposes, and relate it to the previous results. This leads to the discussion of several algorithms to actually perform the theoretical constructions. Finally, we compute the spaces of modular symbols for a couple of subgroups using the Sage software [15], which implements more sophisticated versions of the algorithms explained. The output of Sage is analysed in comparison with the previous exposition in order to exemplify all the important constructions.

The exposition of the first part of this chapter is based mainly on Cremona's book [1]. In turn, the book [14] by Stein has been especially useful for the examples and the concrete computations with Sage. (In fact, Stein is the lead developer of Sage.)

5.1 Alternative presentation of modular symbols

In this section, we follow a different approach to define modular symbols in a way which is more appropriate for computations. This offers a different perspective of the theory of modular symbols and serves as a summary of the results developed in chapter 4.

Definition 5.1. The space $\mathbb{M}$ of *formal modular symbols* is the free abelian group generated by the formal symbols $\{r, s\}$ for $r, s \in \mathbb{P}^1_{\mathbb{Q}}$ modulo the relations

$$\{r, s\} + \{s, t\} + \{t, r\} = 0$$

for all $r, s, t \in \mathbb{P}^1_{\mathbb{Q}}$ and modulo any torsion.

Since we have defined $\mathbb{M}$ to be torsion-free, the relation $\{r,r\}+\{r,r\}+\{r,r\}=0$ implies that $\{r,r\}=0$ for all $r\in\mathbb{P}^1_{\mathbb{Q}}$. Now, from the relation $\{r,r\}+\{r,s\}+\{s,r\}=0$, we deduce that $\{r,s\}+\{s,r\}=0$ for all $r,s\in\mathbb{P}^1_{\mathbb{Q}}$.

The elements of $\mathbb{M}$ can be thought of as formal sums of paths between cusps in $\mathbb{H}^*$, by analogy with the definition of modular symbols given in chapter 4. As a matter of fact, we are considering homotopy classes of paths and so, by lemma 4.6, these are defined uniquely by their endpoints. The relations introduced in the definition of $\mathbb{M}$ correspond to the concatenation of such paths.

Definition 5.2. We define a left action of $\mathrm{GL}_2^+(\mathbb{Q})$ on $\mathbb{M}$ given by

$$\alpha\{r,s\}=\{\alpha(r),\alpha(s)\}$$

for all $r,s\in\mathbb{P}^1_{\mathbb{Q}}$ and all $\alpha\in\mathrm{GL}_2^+(\mathbb{Q})$ and extended linearly.

Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$.

Definition 5.3. The space $\mathbb{M}(\Gamma)$ of (formal) *modular symbols* for Γ is the quotient of $\mathbb{M}$ by the subgroup generated by the elements $\{r,s\}-\gamma\{r,s\}$ for all $r,s\in\mathbb{P}^1_{\mathbb{Q}}$ and all $\gamma\in\Gamma$ and modulo any torsion.

Sometimes, the notation $\{r,s\}_\Gamma$ is used to refer to the image of a (formal) modular symbol $\{r,s\}$ in $\mathbb{M}(\Gamma)$. However, we omit the subscript from the notation as the context will make clear whether we are working with elements of $\mathbb{M}$ or with elements of $\mathbb{M}(\Gamma)$.

The group $\mathbb{M}(\Gamma)$, along with the action of $\mathrm{GL}_2^+(\mathbb{Q})$ induced by the action of the same group on $\mathbb{M}$, satisfies all the properties of proposition 4.9. Hence, the generators of $\mathbb{M}(\Gamma)$ behave essentially in the same way as the modular symbols for Γ presented in definition 4.7.

Definition 5.4. The space $\mathbb{B}(\Gamma)$ of *boundary symbols* for Γ is the free abelian group generated by $\mathrm{C}(\Gamma)$. For each $s\in\mathbb{P}^1_{\mathbb{Q}}$, we write $\{s\}$ for the generator of $\mathbb{B}(\Gamma)$ corresponding to Γs, so $\{s\}=\{\gamma(s)\}$ for all $\gamma\in\Gamma$.

Definition 5.5. The *boundary map* is the morphism

$$\delta\colon \mathbb{M}(\Gamma)\longrightarrow\mathbb{B}(\Gamma)$$

defined by $\delta(\{r,s\})=\{s\}-\{r\}$ for all $\{r,s\}\in\mathbb{M}(\Gamma)$ and extended linearly. The kernel $\mathbb{S}(\Gamma)$ of δ is the subspace of *cuspidal modular symbols* for Γ.

Proposition 5.6. *Let $\overline{\alpha_1}, \ldots, \overline{\alpha_m}$ be a set of representatives of the right cosets of $\overline{\Gamma}$ in* $\mathrm{PSL}_2(\mathbb{Z})$. *Every $\{r,s\} \in \mathbb{M}(\Gamma)$ can be expressed as a $\mathbb{Z}$–linear combination of the form*

$$\sum_{j=1}^{m} \lambda_j \{\alpha_j(0), \alpha_j(\infty)\}.$$

Proof. We can write $\{r,s\} = \{0,s\} - \{0,r\}$ and then proceed exactly as in the proof of theorem 4.12 (using continued fractions). □

Lemma 5.7. *The morphism*

$$\begin{aligned} \psi\colon \mathrm{Man}(\Gamma) &\longrightarrow \mathbb{M}(\Gamma) \\ (\overline{\alpha}) &\longmapsto \{\alpha(0), \alpha(\infty)\} \end{aligned}$$

induces an isomorphism between $\mathrm{C}(\mathrm{Man}(\Gamma))/\mathrm{B}(\mathrm{Man}(\Gamma))$ *and* $\mathbb{M}(\Gamma)$ *and also an isomorphism between* $\mathrm{Z}(\mathrm{Man}(\Gamma))/\mathrm{B}(\mathrm{Man}(\Gamma))$ *and* $\mathbb{S}(\Gamma)$.

Proof. First, we observe that ψ is well-defined. Let $(\overline{\alpha})$ be a Manin symbol. Since in $\mathbb{M}(\Gamma)$ we have that $\{\alpha(0), \alpha(\infty)\} = \{\pm\gamma\alpha(0), \pm\gamma\alpha(\infty)\}$ for all $\gamma \in \Gamma$, the modular symbol $\psi((\overline{\alpha})) = \{\alpha(0), \alpha(\infty)\}$ does not depend on the representative α of the right coset $\overline{\Gamma}\,\overline{\alpha}$. Furthermore, proposition 5.6 implies that ψ is surjective.

On the one hand, the quotient $\mathrm{C}(\mathrm{Man}(\Gamma))/\mathrm{B}(\mathrm{Man}(\Gamma))$ is $\mathrm{Man}(\Gamma)$ modulo the relations $(\overline{\alpha}) + (\overline{\alpha\sigma}) = 0$, $(\overline{\alpha}) = 0$ if $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\sigma}$, $(\overline{\alpha}) + (\overline{\alpha\tau}) + (\overline{\alpha\tau^2}) = 0$ and $(\overline{\alpha}) = 0$ if $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\tau}$. On the other hand, the analogous relations $\alpha\{0,\infty\} + \alpha\sigma\{0,\infty\} = 0$, $\alpha\{0,\infty\} = 0$ if $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\sigma}$, $\alpha\{0,\infty\} + \alpha\tau\{0,\infty\} + \alpha\tau^2\{0,\infty\} = 0$ and $\alpha\{0,\infty\} = 0$ if $\overline{\Gamma}\,\overline{\alpha} = \overline{\Gamma}\,\overline{\alpha\tau}$ hold in $\mathbb{M}(\Gamma)$ (see the beginning of section 4.3). Therefore, ψ induces an epimorphism $\widehat{\psi}\colon \mathrm{C}(\mathrm{Man}(\Gamma))/\mathrm{B}(\mathrm{Man}(\Gamma)) \to \mathbb{M}(\Gamma)$. Similarly, ψ induces an epimorphism $\widetilde{\psi}\colon \mathrm{Z}(\mathrm{Man}(\Gamma))/\mathrm{B}(\mathrm{Man}(\Gamma)) \to \mathbb{S}(\Gamma)$ because ψ preserves the boundaries, in the sense that $\delta(\psi((\overline{\alpha}))) = \partial(\overline{\alpha})$ for every Manin symbol $(\overline{\alpha})$.

Recall that in the last part of the proof of theorem 4.19 we used a monomorphism $\varphi\colon \mathrm{C}(\mathrm{Man}(\Gamma)) \to \mathrm{C}_1(\mathrm{L})$. We proved that $\varphi(\mathrm{B}(\mathrm{Man}(\Gamma))) = \mathrm{B}_1(\mathrm{L})$ and so φ induces a monomorphism $\widehat{\varphi}\colon \mathrm{C}(\mathrm{Man}(\Gamma))/\mathrm{B}(\mathrm{Man}(\Gamma)) \to \mathrm{C}_1(\mathrm{L})/\mathrm{B}_1(\mathrm{L})$. And, for every Manin symbol $(\overline{\alpha})$, $\widehat{\varphi}((\overline{\alpha}) + \mathrm{B}(\mathrm{Man}(\Gamma)))$ coincides with the homology class of the projection in $\mathrm{X}(\Gamma)$ of a path from $\alpha(0)$ to $\alpha(\infty)$, which was represented as $\{\alpha(0), \alpha(\infty)\}$ in chapter 4. Since those modular symbols (defined in terms of the homology of $\mathrm{X}(\Gamma)$) satisfy all the relations which are satisfied by the elements of $\mathbb{M}(\Gamma)$ (by proposition 4.9), $\widehat{\psi}$ factors over $\widehat{\varphi}$. That is, we can express $\widehat{\varphi} = f \circ \widehat{\psi}$, where f is the natural morphism which maps (formal) modular symbols to

homology classes of the corresponding paths. Therefore, $\widehat{\psi}$ must be injective. As a consequence, the restriction $\widetilde{\psi}$ of $\widehat{\psi}$ is also injective. □

Theorem 5.8. *The natural morphism*

$$\phi\colon \mathbb{M}(\Gamma) \longrightarrow H_1(X(\Gamma), C(\Gamma), \mathbb{Z}),$$

which maps a (formal) modular symbol $\{r,s\}$ to the homology class $\{r,s\}$ (in the sense of definition 4.7) of the projection in $X(\Gamma)$ of a path from r to s in $\mathbb{H}^*$, is an isomorphism. Moreover, ϕ induces a canonical isomorphism ξ' between $\mathbb{S}(\Gamma)$ and $H_1(X(\Gamma), \mathbb{Z})$.

Proof. As we observed in the proof of lemma 5.7, proposition 4.9 implies that ϕ is well-defined.

First we are going to prove that ξ' is an isomorphism using theorem 4.19. Then, we are going to use this fact in order to prove that ϕ is also an isomorphism.

Let $\widehat{\psi}\colon C(\mathrm{Man}(\Gamma))/B(\mathrm{Man}(\Gamma)) \to \mathbb{M}(\Gamma)$ and $\widetilde{\psi}\colon Z(\mathrm{Man}(\Gamma))/B(\mathrm{Man}(\Gamma)) \to \mathbb{S}(\Gamma)$ be the two isomorphisms described in lemma 5.7. The composition $\xi' \circ \widetilde{\psi}$ is precisely the isomorphism $\xi\colon Z(\mathrm{Man}(\Gamma))/B(\mathrm{Man}(\Gamma)) \to H_1(X(\Gamma), \mathbb{Z})$ described in theorem 4.19. Therefore, ξ' is an isomorphism.

Recall that the relative homology is the homology of a quotient of chain complexes and so there is a long exact sequence

$$\begin{aligned} \cdots \longrightarrow H_p(C(\Gamma), \mathbb{Z}) \longrightarrow H_p(X(\Gamma), \mathbb{Z}) \longrightarrow H_p(X(\Gamma), C(\Gamma), \mathbb{Z}) \longrightarrow \\ \longrightarrow H_{p-1}(C(\Gamma), \mathbb{Z}) \longrightarrow \cdots \qquad \cdots \longrightarrow H_0(X(\Gamma), C(\Gamma), \mathbb{Z}) \longrightarrow 0 \end{aligned}$$

induced by the short exact sequence of chain complexes. We are only interested in the last part of the long exact sequence. Since $C(\Gamma)$ is a finite set of points (with the discrete topology), $H_1(C(\Gamma), \mathbb{Z}) = 0$ and $H_0(C(\Gamma), \mathbb{Z}) \cong \mathbb{Z}^{|C(\Gamma)|}$. In addition, $H_0(X(\Gamma), \mathbb{Z}) \cong \mathbb{Z}$ and $H_0(X(\Gamma), C(\Gamma), \mathbb{Z}) = 0$ because $X(\Gamma)$ is path-connected and $C(\Gamma) \neq \emptyset$. Hence, there is an exact sequence

$$0 \longrightarrow H_1(X(\Gamma), \mathbb{Z}) \xrightarrow{i_1} H_1(X(\Gamma), C(\Gamma), \mathbb{Z}) \xrightarrow{\Delta} \mathbb{Z}^{|C(\Gamma)|} \xrightarrow{\Sigma} \mathbb{Z} \longrightarrow 0$$

given by the previous long exact sequence. Here, i_1 is the morphism induced by the inclusion of $C(\Gamma)$ in $X(\Gamma)$. The connecting morphism Δ maps the homology class of a closed path or a path with endpoints in $C(\Gamma)$ to its boundary; we identify $H_0(C(\Gamma), \mathbb{Z})$ (which is the free abelian group generated by $C(\Gamma)$) with $\mathbb{Z}^{|C(\Gamma)|}$. Finally, the morphism Σ maps a formal $\mathbb{Z}$–linear combination of the

elements of $C(\Gamma)$ to the sum of its coefficients (i.e., its degree) because this corresponds to the morphism induced by the inclusion of $C(\Gamma)$ in $X(\Gamma)$ (all the points of $X(\Gamma)$ are equivalent in $H_0(X(\Gamma),\mathbb{Z})$).

In view of our definition of $\mathbb{S}(\Gamma)$, we can define an exact sequence

$$0 \longrightarrow \mathbb{S}(\Gamma) \xrightarrow{i} \mathbb{M}(\Gamma) \xrightarrow{\delta} \mathbb{B}(\Gamma) \xrightarrow{\deg} \mathbb{Z} \longrightarrow 0$$

resembling the previous one. Here, i is the inclusion morphism and deg is the morphism which maps a formal $\mathbb{Z}$–linear combination of the elements of $C(\Gamma)$ to the sum of its coefficients (i.e., its degree). We only need to check that deg is surjective and that the image of δ is precisely the kernel of deg. On the one hand, for each $n \in \mathbb{Z}$, $\deg(n\{\infty\}) = n$. Thus, deg is surjective. On the other hand, for every modular symbol $\{r,s\}$, $\deg(\delta(\{r,s\})) = \deg(\{s\}-\{r\}) = 1-1 = 0$. Conversely, an element of the kernel of deg is of the form $\{s_1\}+\cdots+\{s_n\}-\{r_1\}-\cdots-\{r_n\}$ and this is the image of $\{r_1,s_1\}+\cdots+\{r_n,s_n\}$ under δ.

Actually, those two exact sequences are roughly the same. Specifically, the diagram

$$\begin{array}{ccccccccccc}
0 & \longrightarrow & \mathbb{S}(\Gamma) & \xrightarrow{i} & \mathbb{M}(\Gamma) & \xrightarrow{\delta} & \mathbb{B}(\Gamma) & \xrightarrow{\deg} & \mathbb{Z} & \longrightarrow & 0 \\
\| & & \downarrow{\scriptstyle \xi'} & & \downarrow{\scriptstyle \phi} & & \uparrow{\scriptstyle \wr} & & \| & & \| \\
0 & \longrightarrow & H_1(X(\Gamma),\mathbb{Z}) & \xrightarrow{i_1} & H_1(X(\Gamma),C(\Gamma),\mathbb{Z}) & \xrightarrow{\Delta} & \mathbb{Z}^{|C(\Gamma)|} & \xrightarrow{\Sigma} & \mathbb{Z} & \longrightarrow & 0
\end{array}$$

is commutative and has exact rows. Moreover, the morphisms in the columns are all isomorphisms except for possibly ϕ. Therefore, by the five lemma (from homological algebra), we conclude that ϕ must be an isomorphism too. □

As mentioned before, this presentation is very suitable for computations. For instance, many algorithms to compute the short exact sequence of modular symbols appearing in the proof of theorem 5.8 are implemented in Sage. The importance of modular symbols lies in part in the fact that $\mathbb{S}(\Gamma)$ is a module on which the Hecke algebra acts in a quite simple way. Thus, one can study properties of the Hecke operators by means of their action on modular symbols and then translate the results to the theory of modular forms.

From now on, we fix $N \in \mathbb{N}$ and focus on the case in which $\Gamma = \Gamma_0(N)$, so that we can use the results of chapter 3 as well.

Definition 5.9. Let $n \in \mathbb{N}$. We define the action of the n–th *Hecke operator* on $\mathbb{M}(\Gamma_0(N))$, $T(n)\colon \mathbb{M}(\Gamma_0(N)) \to \mathbb{M}(\Gamma_0(N))$, as follows: for all $\{r,s\} \in \mathbb{M}(\Gamma_0(N))$,

$$T(n)\{r,s\} = \sum_{a,b,d} \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \{r,s\},$$

where the sum is over the triples of integers a, b and d such that $a \geq 1$, $(a,N) = 1$, $ad = n$ and $0 \leq b < d$, and this is extended by linearity.

We observe that this definition of Hecke operators on (formal) modular symbols coincides with definition 4.5 (up to the isomorphism described in theorem 5.8). Let $f \in S_2(\Gamma_0(N))$ and let $r, s \in \mathbb{P}^1_{\mathbb{Q}}$. By theorem 5.8, we have that

$$T(n) f(z) = \sum_{a,b,d} \frac{a}{d} f\left(\frac{az+b}{d}\right).$$

On the other hand, for every $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$,

$$\int_r^s f|_2^{[\alpha]}\, dz = \int_r^s f(\alpha(z))\, d\alpha(z) = \int_{\alpha(r)}^{\alpha(s)} f(z)\, dz.$$

That is why the action of $T(n)$ on modular symbols is defined using the same matrices as the action of $T(n)$ on modular forms. Consequently, $T(n)$ acts on $\mathbb{S}(\Gamma_0(N))$ (because it acts on $S_2(\Gamma_0(N))$).

It is often useful to combine modular symbols and Manin symbols. On the one hand, by proposition 4.21, every Manin symbol for $\Gamma_0(N)$ can be identified with an element $(c : d)$ of $\mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$. On the other hand, lemma 5.7 gives us an explicit isomorphism between $\mathbb{M}(\Gamma_0(N))$ and $C(\mathrm{Man}(\Gamma_0(N)))/B(\mathrm{Man}(\Gamma_0(N)))$. Combining these results, we can identify $\mathbb{M}(\Gamma_0(N))$ with the free abelian group generated by $\mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$ modulo the relations

$$(c : d) + (-d : c) = 0 \quad \text{and} \quad (c : d) + (c + d : -c) + (d : -c - d) = 0$$

for all $(c : d) \in \mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$ (these are the relations given by the matrices σ and τ) and modulo any torsion.

With this interpretation, the boundary of a Manin symbol $(c : d)$ can be computed as follows: we build a matrix $\left(\begin{smallmatrix} a & b \\ c_0 & d_0 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that $(c : d) = (c_0 : d_0)$ as in the proof of proposition 4.21 and then $\delta((c : d)) = \left\{\frac{a}{c_0}\right\} - \left\{\frac{b}{d_0}\right\}$. To compute δ, we need a criterion to know whether two cusps are $\Gamma_0(N)$–equivalent.

Lemma 5.10. *For $j \in \{1,2\}$, let $v_j = \frac{p_j}{q_j}$, where $p_j, q_j \in \mathbb{Z}$ with $(p_j, q_j) = 1$. The following conditions are equivalent:*

(a) *$v_2 = \gamma(v_1)$ for some $\gamma \in \Gamma_0(\mathrm{N})$;*

(b) *there exists $\lambda \in \mathbb{Z}$ with $(\lambda, \mathrm{N}) = 1$ such that $q_2 \equiv \lambda q_1$ and $\lambda p_2 \equiv p_1 \pmod{\mathrm{N}}$;*

(c) *there exist $s_1, s_2 \in \mathbb{Z}$ satisfying that $s_1 p_1 \equiv 1 \pmod{q_1}$, $s_2 p_2 \equiv 1 \pmod{q_2}$ and $s_1 q_2 \equiv s_2 q_1 \pmod{(q_1 q_2, \mathrm{N})}$.*

Proof. First we check that (a) $\Longrightarrow$ (b). Suppose that $\gamma = \left(\begin{smallmatrix} a & b \\ \mathrm{N}c & d \end{smallmatrix}\right)$ and $\gamma(v_1) = v_2$. Since $(p_1, q_1) = 1$ and γ is invertible,

$$v_2 = \gamma(v_1) = \frac{ap_1 + bq_1}{\mathrm{N}cp_1 + dq_1}$$

is written in lowest terms. Hence, $p_2 = \pm(ap_1 + bq_1)$ and $q_2 = \pm(\mathrm{N}cp_1 + dq_1)$. In particular, we can choose $\lambda = \pm d$ so that $\lambda q_1 \equiv q_2$ and $\lambda p_2 \equiv p_1 \pmod{\mathrm{N}}$.

Now we prove that (b) $\Longrightarrow$ (a). Let $\mathrm{D} = (q_1, \mathrm{N}) = (q_2, \mathrm{N})$ (the two greatest common divisors are the same because $q_2 \equiv \lambda q_1 \pmod{\mathrm{N}}$). Consider $x_1, y_1, x_2, y_2 \in \mathbb{Z}$ such that $x_1 p_1 - y_1 q_1 = x_2 p_2 - y_2 q_2 = 1$ (Bézout's identity). By hypothesis, we have that $\lambda p_2 \equiv p_1 \pmod{\mathrm{D}}$, so $\lambda x_1 \equiv \lambda p_2 x_2 x_1 \equiv p_1 x_1 x_2 \equiv x_2 \pmod{\mathrm{D}}$. And, since $(\lambda, \mathrm{N}) = 1$, the equation $\frac{\lambda q_1}{\mathrm{D}}\mathrm{X} \equiv \frac{\lambda x_1 - x_2}{\mathrm{D}} \pmod{\frac{\mathrm{N}}{\mathrm{D}}}$ has a solution, which implies that there exists $\mu \in \mathbb{Z}$ such that $\mu\lambda q_1 \equiv \lambda x_1 - x_2 \pmod{\mathrm{N}}$. Define $s_1 = x_1 - \mu q_1$, $r_1 = y_1 - \mu p_1$, $s_2 = x_2$ and $r_2 = y_2$, so that $p_j s_j - q_j r_j = 1$ for $j \in \{1,2\}$. In particular, $\left(\begin{smallmatrix} p_j & r_j \\ q_j & s_j \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Since $\lambda q_1 \equiv q_2$ and $\lambda s_1 \equiv s_2 \pmod{\mathrm{N}}$, lemma 4.20 asserts that there exists a matrix $\gamma \in \Gamma_0(\mathrm{N})$ such that $\left(\begin{smallmatrix} p_2 & r_2 \\ q_2 & s_2 \end{smallmatrix}\right) = \gamma\left(\begin{smallmatrix} p_1 & r_1 \\ q_1 & s_1 \end{smallmatrix}\right)$. Furthermore, observe that $\left(\begin{smallmatrix} p_j & r_j \\ q_j & s_j \end{smallmatrix}\right)\infty = v_j$ for $j \in \{1,2\}$. Consequently, $v_2 = \gamma(v_1)$.

Let us prove that (a) $\Longrightarrow$ (c). As before, we can choose $r_1, r_2, s_1, s_2 \in \mathbb{Z}$ such that $p_1 s_1 - q_1 r_1 = p_2 s_2 - q_2 r_2 = 1$ and $\gamma = \alpha_2 \alpha_1^{-1} \in \Gamma_0(\mathrm{N})$, where $\alpha_j = \left(\begin{smallmatrix} p_j & r_j \\ q_j & s_j \end{smallmatrix}\right)$. One checks easily that $\gamma \in \Gamma_0(\mathrm{N})$ if and only if $q_2 s_1 - q_1 s_2 \equiv 0 \pmod{\mathrm{N}}$.

Finally, we have to prove that (c) $\Longrightarrow$ (a). Again, consider $r_1, r_2 \in \mathbb{Z}$ such that $p_1 s_1 - q_1 r_1 = p_2 s_2 - q_2 r_2 = 1$. Define $\alpha_j = \left(\begin{smallmatrix} p_j & r_j \\ q_j & s_j \end{smallmatrix}\right)$ for $j \in \{1,2\}$ and $\gamma = \alpha_2 \alpha_1^{-1}$, which satisfies that $\gamma(v_1) = v_2$. As before, $\gamma \in \Gamma_0(\mathrm{N})$ if and only if $q_2 s_1 - q_1 s_2 \equiv 0 \pmod{\mathrm{N}}$. This is not necessarily the case. But we can replace s_1 with a number of the form $s_1' = s_1 + x q_1$ (and, similarly, r_1 with $r_1' = r_1 + x p_1$) and obtain thus a matrix γ' with the same properties. In particular, $\gamma' \in \Gamma_0(\mathrm{N})$ if and only if $q_2 s_1 - q_1 s_2 + x q_1 q_2 \equiv 0 \pmod{\mathrm{N}}$. And the equation $q_1 q_2 \mathrm{X} \equiv q_1 s_2 - q_2 s_1 \pmod{\mathrm{N}}$ has a solution: if $\mathrm{D} = (q_1 q_2, \mathrm{N})$, $q_1 s_2 - q_2 s_1 \equiv 0 \pmod{\mathrm{D}}$ (by hypothesis) and the equation $\frac{q_1 q_2}{\mathrm{D}}\mathrm{X} \equiv \frac{q_1 s_2 - q_2 s_1}{\mathrm{D}} \pmod{\frac{\mathrm{N}}{\mathrm{D}}}$ has a solution. □

This technical lemma provides a way to verify if an element of $\mathbb{M}(\Gamma_0(N))$ belongs to $\mathbb{S}(\Gamma_0(N))$.

5.2 Ideas for the algorithms

This section explains some algorithms which are obtained adapting the proofs of the previous results. The main objective of the section is neither to provide a fully detailed description of these algorithms nor to study the implementation details in depth, but to convey the idea that the constructions described before can actually be implemented in a computer. In particular, we do not strive for obtaining the best known algorithms and the algorithms are explained in a quite informal way (with no pseudocode).

Conversions between modular symbols and Manin symbols. As explained at the end of the previous section, lemma 5.7 allows us to identify the elements of $\mathbb{M}(\Gamma_0(N))$ with the elements of the free abelian group generated by $\mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$ modulo the relations $(c:d)+(-d:c)=0$ and $(c:d)+(c+d:-c)+(d:-c-d)=0$ for all $(c:d)\in\mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$ and modulo any torsion. What is more, the conversions between the two representations (i.e., using modular symbols or using Manin symbols) can be performed algorithmically.

On the one hand, an element of $\mathbb{M}(\Gamma_0(N))$ is a $\mathbb{Z}$–linear combination of modular symbols $\{r,s\}$ with $r,s\in\mathbb{P}^1_{\mathbb{Q}}$. By proposition 5.6, each modular symbol $\{r,s\}\in\mathbb{M}(\Gamma_0(N))$ can be expressed as a $\mathbb{Z}$–linear combination of modular symbols of the form $\alpha\{0,\infty\}$ with $\alpha\in\mathrm{SL}_2(\mathbb{Z})$: to do so, one only has to write $\{r,s\}=\{0,s\}-\{0,r\}$ and then compute the successive convergents of the (finite) continued fraction representations of r and s in order to use Manin's trick, as in the proof of theorem 4.12. Finally, if $\alpha=\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$, the modular symbol $\alpha\{0,\infty\}$ corresponds to the Manin symbol $(c:d)$, by lemma 5.7 and proposition 4.21.

On the other hand, given a Manin symbol $(c:d)\in\mathbb{P}^1_{\mathbb{Z}/N\mathbb{Z}}$, we can find a matrix $\alpha=\left(\begin{smallmatrix}a & b\\ c_0 & d_0\end{smallmatrix}\right)\in\mathrm{SL}_2(\mathbb{Z})$ such that $(c_0:d_0)=(c:d)$ by computing a greatest common divisor and a pair of Bézout coefficients, as in the proof of proposition 4.21. In this case, lemma 5.7 implies that the Manin symbol $(c:d)$ corresponds to the modular symbol $\left\{\frac{b}{d_0},\frac{a}{c_0}\right\}$.

Computing a basis of $\mathbb{M}(\Gamma_0(N))$. The space $\mathbb{M}(\Gamma_0(N))$ is a free $\mathbb{Z}$–module and so it has a basis, which we can compute using Manin symbols.

Lemma 4.20 gives a criterion to determine whether two Manin symbols are the same. Using it, we can make a list of inequivalent Manin symbols for $\Gamma_0(N)$ as follows. First, we list the symbols $(c : 1)$ with $0 \leq c < N$; then, the symbols $(1 : d)$ with $0 \leq d < N$ and $(d, N) > 1$; finally, a set of pairwise inequivalent symbols $(c : d)$ with $c \mid N$, $c \notin \{1, N\}$, $(c, d) = 1$ and $(d, N) > 1$. By condition (b) of lemma 4.20, it is clear that these Manin symbols are all inequivalent. Moreover, one checks that every Manin symbol is of one of these forms. Indeed, every Manin symbol has a representative $(c_0 : d_0)$ with $(c_0, d_0) = 1$. If $(d_0, N) = 1$, there is a multiplicative inverse d_0^{-1} of $d_0 \pmod N$ and so $(c_0 : d_0) = (d_0^{-1} c_0 : 1)$. If $(d_0, N) \neq 1$ but $(c_0, N) = 1$, there is a multiplicative inverse c_0^{-1} of $c_0 \pmod N$ and so $(c_0 : d_0) = (1 : c_0^{-1} d_0)$. Otherwise, we consider $c = (c_0, N)$ and write $c_0 = cx$ and $N = cn$. Since $(x, n) = 1$, there exists $y \in \mathbb{Z}$ such that $xy \equiv 1 \pmod n$; in fact, we can choose $\lambda \in \mathbb{Z}$ such that $\lambda x \equiv 1 \pmod n$ and $(\lambda, N) = 1$ (because the projection $(\mathbb{Z}/N\mathbb{Z})^\times \twoheadrightarrow (\mathbb{Z}/n\mathbb{Z})^\times$ is surjective). Therefore, we can define $d = \lambda d_0$ and we have that $(c_0 : d_0) = (\lambda c_0 : \lambda d_0) = (c : d)$ with $c \mid N$, $c \notin \{1, N\}$, $(c, d) = 1$ and $(d, N) > 1$.

Having a list of inequivalent Manin symbols, we take into account the relations of the forms $(c : d) + (-d : c) = 0$ and $(c : d) + (c + d : -c) + (d : -c - d) = 0$ in order to obtain a basis from this (finite) set of generators. For instance, this can be done by representing these relations as the columns of a matrix whose rows are indexed by the generators and then computing the Smith normal form of that matrix (there is an algorithm, which is very similar to Gaussian elimination, to compute the Smith normal form of a matrix).

Computing a basis of $\mathbb{S}(\Gamma_0(N))$. The space $\mathbb{S}(\Gamma_0(N))$ is the kernel of the boundary map $\delta\colon \mathbb{M}(\Gamma_0(N)) \to \mathbb{B}(\Gamma_0(N))$. Hence, we need to compute the images of the elements of the previously obtained basis of $\mathbb{M}(\Gamma_0(N))$ under δ.

Recall that lemma 5.10 gives a criterion to determine whether two cusps are $\Gamma_0(N)$–equivalent. This is enough to compute the boundary map without computing a basis of $\mathbb{B}(\Gamma_0(N))$ in advance. Instead, we can keep a cumulative list of the cusps found so far, so that each cusp encountered while computing δ is checked for equivalence with those already in the list and is added to the list if it corresponds to a new equivalence class of cusps.

Using the previous trick to compute the $\Gamma_0(N)$–equivalence classes of cusps while computing the image of a basis of $\mathbb{M}(\Gamma_0(N))$ under δ, we can compute a matrix with integer entries for the linear map δ. Then, we can compute its Hermite normal form (with an algorithm which is very similar to Gaussian

elimination) and obtain thus a basis of $\mathbb{S}(\Gamma_0(N))$.

Computing a basis of $S_2(\Gamma_0(N))$ using Hecke operators. Proposition 4.2 gives an isomorphism of complex vector spaces between $S_2(\Gamma_0(N))$ and $\mathrm{Hom}_{\mathbb{C}}(\mathbb{T}_{\mathbb{C}}, \mathbb{C})$, where $\mathbb{T}_{\mathbb{C}}$ is the complex vector space generated by the Hecke operators acting on $S_2(\Gamma_0(N))$. We can use this isomorphism to compute a basis of $S_2(\Gamma_0(N))$. Moreover, proposition 4.3 provides a method to obtain the q–expansion of a cusp form from the corresponding element of $\mathrm{Hom}_{\mathbb{C}}(\mathbb{T}_{\mathbb{C}}, \mathbb{C})$.

Hecke operators can be regarded as linear maps acting on $\mathbb{S}(\Gamma_0(N))$, which is a free $\mathbb{Z}$–module of rank $2g$ (where g is the genus of $X_0(N)$). Therefore, we can compute a matrix giving the action of any Hecke operator with respect to the previously computed basis of $\mathbb{S}(\Gamma_0(N))$. To do so, we express an element of the basis (computed using Manin symbols) as a sum of modular symbols, compute the action of the Hecke operator on each of these modular symbols by definition and express the resulting modular symbols in terms of Manin symbols. With this procedure, we obtain the $2g \times 2g$ matrix of the Hecke operator.

There are alternative approaches to compute these matrices directly with Manin symbols. For instance, Cremona's book [1] explains how to use what are known as Heilbronn matrices to compute the action of Hecke operators on Manin symbols and describes an algorithm to compute them: for each prime p such that $p \nmid N$, there is a set of Heilbronn matrices of level p which act (on the right) on a Manin symbol in the same way as the Hecke operator $T(p)$ acts on the corresponding modular symbol. The proof of the fact that these Heilbronn matrices correspond to the action of the Hecke operators is adapted from Merel's paper [6], where Heilbronn matrices are introduced in a more general setting. In his paper [6], Merel also presents several other families of matrices which are useful in more general contexts.

Since the action of a Hecke operator $T(n)$ on $S_2(\Gamma_0(N))$ is dual to the action of $T(n)$ on $\mathbb{S}(\Gamma_0(N)) \otimes_{\mathbb{Z}} \mathbb{R}$ (up to the isomorphisms described in theorem 4.4 and in theorem 5.8), we can use the matrices of Hecke operators acting on $\mathbb{S}(\Gamma_0(N))$, obtained with the procedure explained before, to describe $\mathbb{T}_{\mathbb{C}}$.

Let $[T(n)]$ be the $2g \times 2g$ matrix associated with $T(n)$ (with respect to the previously computed basis of $\mathbb{S}(\Gamma_0(N))$). The operator $[\,\cdot\,]$ defines a linear map which embeds $\mathbb{T}_{\mathbb{C}}$ in the space of $2g \times 2g$ matrices. If A is a $2g \times 2g$ matrix, we write $a_{ij}(A)$ for the entry in the i–th row and the j–th column of A for all $i, j \in \{1, \ldots, 2g\}$. Since the linear maps $a_{ij}(\cdot)$ for $i, j \in \{1, \ldots, 2g\}$ form a basis of the

dual space of the vector space of $2g \times 2g$ matrices and we can view $\mathrm{Hom}_{\mathbb{C}}(\mathbb{T}_{\mathbb{C}}, \mathbb{C})$ as a subspace of this dual space, the induced linear maps $a_{ij}(\cdot)$ for $i, j \in \{1, \ldots, 2g\}$ defined on $\mathbb{T}_{\mathbb{C}}$ (by restriction) generate $\mathrm{Hom}_{\mathbb{C}}(\mathbb{T}_{\mathbb{C}}, \mathbb{C})$.

In conclusion, we can compute the q–expansions of the elements of a basis of $\mathrm{S}_2(\Gamma_0(\mathrm{N}))$ to precision $\mathrm{O}(q^{\mathrm{B}+1})$ for $\mathrm{B} \in \mathbb{N}$ as follows. By proposition 4.3, we can define cusp forms f_{ij} given by the q–expansions

$$\widehat{(f_{ij})}_{\infty}(q) = \sum_{n=1}^{\mathrm{B}} a_{ij}([\mathrm{T}(n)])q^n + \mathrm{O}(q^{\mathrm{B}+1})$$

for all $i, j \in \{1, \ldots, 2g\}$. The arguments exposed in the previous paragraph imply that these f_{ij} for $i, j \in \{1, \ldots, 2g\}$ generate $\mathrm{S}_2(\Gamma_0(\mathrm{N}))$. Hence, we only need to choose g linearly independent cusp forms amongst these. We can use Gaussian elimination with the first B coefficients of the q–expansions of these cusp forms in order to obtain a basis (if B is large enough, we are going to be able to distinguish when two of these cusp forms are distinct). Observe that, with this algorithm, we obtain a basis of cusp forms of weight 2 for $\Gamma_0(\mathrm{N})$ with integral Fourier coefficients: this is just one example of a non-trivial result which can be obtained as a direct consequence of the results from the theory of modular symbols. Similarly, using the fact that the matrices $[\mathrm{T}(n)]$ for $n \in \mathbb{N}$ have integer entries, we deduce that the eigenvalues of the Hecke operators are algebraic integers and that the q–expansions of the normalised Hecke eigenforms in $\mathrm{S}_2(\Gamma_0(\mathrm{N}))$ are algebraic over $\mathbb{Q}$.

Theorem 5.11. *There exists a basis of* $\mathrm{S}_2(\Gamma_0(\mathrm{N}))$ *consisting of forms whose Fourier coefficients are integers.*

Theorem 5.12. *For every* $n \in \mathbb{N}$*, the eigenvalues of the Hecke operator* $\mathrm{T}(n)$ *(regarded as an endomorphism of* $\mathrm{S}_2(\Gamma_0(\mathrm{N}))$*) are algebraic integers.*

As a matter of fact, one could obtain a basis of $\mathrm{S}_2(\Gamma_0(\mathrm{N}))$ consisting of Hecke eigenforms with similar methods using modular symbols. In this case, though, one needs to combine the theory of modular symbols with the theory of oldforms and newforms (also known as Atkin–Lehner theory). One defines what are known as degeneracy maps between $\mathbb{S}(\Gamma_0(\mathrm{N}))$ and $\mathbb{S}(\Gamma_0(\mathrm{M}))$ for the divisors M of N and uses them to identify the subspace of $\mathbb{S}(\Gamma_0(\mathrm{N}))$ formed of cuspidal modular symbols arising from lower levels (i.e., the subspace corresponding to oldforms). We do not go into further detail here.

5.3 Modular symbols for $\Gamma_0(23)$

We exemplify the algorithms and computations explained before for $\Gamma_0(23)$ using Sage [15]. This is a quite simple example because, as we are going to see later, the genus of $X_0(23)$ is 2 and the dimension of all the relevant spaces is small.

We define the space $\mathbb{M}(\Gamma_0(23))$ in Sage as follows:

```
sage: G=Gamma0(23) ; G
Congruence Subgroup Gamma0(23)
sage: M=ModularSymbols(G) ; M
Modular Symbols space of dimension 5 for
   Gamma_0(23) of weight 2 with sign 0 over
   Rational Field
```

In fact, Sage works with the rational vector space $\mathbb{M}(\Gamma_0(23)) \otimes_{\mathbb{Z}} \mathbb{Q}$ by default. We can compute the Manin symbols for $\Gamma_0(23)$ (that is, the elements of $\mathbb{P}^1_{\mathbb{Z}/23\mathbb{Z}}$) with the following command:

```
sage: M.manin_generators()
[(0,1), (1,0), (1,1), (1,2), (1,3), (1,4),
   (1,5), (1,6), (1,7), (1,8), (1,9), (1,10),
   (1,11), (1,12), (1,13), (1,14), (1,15),
   (1,16), (1,17), (1,18), (1,19), (1,20),
   (1,21), (1,22)]
```

Sage does not use the set of representatives which we described in the previous section, but another one which is analogous. In this case, there are no Manin symbols $(c : d)$ with $c \neq 1$ and $d \neq 1$ because 23 is prime.

We can express a Manin symbol as a modular symbol and vice versa as explained in the previous section. We reproduce those computations for a concrete case:

```
sage: mansymb=M.manin_generators()[10]
sage: mansymb.lift_to_sl2z()
[0, -1, 1, 9]
sage: mansymb.modular_symbol_rep()
{-1/9, 0}
sage:
   modsymb=sage.modular.modsym.modular_symbols.
   ModularSymbol(M,0,0,2/5) ; modsymb
```

```
{0, 2/5}
sage: convergents(2/5)
[0, 1/2, 2/5]
sage: modsymb.manin_symbol_rep()
(-5,2) + (2,1)
```

The first part shows that the Manin symbol $(1:9)$ corresponds to the right coset $\Gamma_0(23)\left(\begin{smallmatrix}0 & -1\\ 1 & 9\end{smallmatrix}\right)$ and so to the modular symbol $\left\{-\frac{1}{9}, 0\right\}$. Analogously, the computation of the second part implies (using Manin's trick) that the modular symbol $\left\{0, \frac{2}{5}\right\}$ can be expressed as

$$\left\{0, \frac{1}{2}\right\} + \left\{\frac{1}{2}, \frac{2}{5}\right\} = \begin{pmatrix}1 & 0\\ 2 & 1\end{pmatrix}\{0, \infty\} + \begin{pmatrix}-2 & 1\\ -5 & 2\end{pmatrix}\{0, \infty\}$$

and so corresponds to $(2:1)+(-5:2)$.

Nevertheless, we are only interested in expressing the elements of $\mathbb{M}(\Gamma_0(23))$ in terms of a fixed basis. The command `M.manin_basis` returns a list of indices of the Manin symbols which form a basis of $\mathbb{M}(\Gamma_0(23)) \otimes_{\mathbb{Z}} \mathbb{Q}$. Alternatively, one can ask Sage to compute the basis directly.

```
sage: M.manin_basis()
[1, 18, 20, 21, 22]
sage: [M.manin_generators()[i] for i in
   M.manin_basis()]
[(1,0), (1,17), (1,19), (1,20), (1,21)]
sage: M.basis()
((1,0), (1,17), (1,19), (1,20), (1,21))
```

We can also switch between Manin symbols and modular symbols and express a Manin symbol in terms of the elements of the basis:

```
sage: set_modsym_print_mode('modular')
sage: M.basis()
({Infinity, 0}, {-1/17, 0}, {-1/19, 0}, {-1/20,
   0}, {-1/21, 0})
sage: M((2,5))
-{-1/19, 0} + {-1/20, 0} - {-1/21, 0}
sage: set_modsym_print_mode('manin')
sage: M.basis()
((1,0), (1,17), (1,19), (1,20), (1,21))
```

```
sage: M((2,5))
-(1,19) + (1,20) - (1,21)
```

We can express an element of $\mathbb{M}(\Gamma_0(23))$ as a linear combination of the elements of the computed basis. Conversely, given a vector, we can compute easily the corresponding element of $\mathbb{M}(\Gamma_0(23))$ in terms of Manin symbols.

```
sage: u=(4,2,1,0,5)
sage: z=M(sum(u[i]*M.basis()[i] for i in
   xrange(5))) ; z
4*(1,0) + 2*(1,17) + (1,19) + 5*(1,21)
sage: M.coordinate_vector(z)
(4, 2, 1, 0, 5)
```

To find a basis of $\mathbb{S}(\Gamma_0(23))$, we need to compute the boundary map δ (and, simultaneously, compute $C_0(23)$ in order to study the image of δ in $\mathbb{B}(\Gamma_0(23))$, as explained in the previous section).

```
sage: M.boundary_map()
Hecke module morphism boundary map defined by
   the matrix
[ 1 -1]
[ 0  0]
[ 0  0]
[ 0  0]
[ 0  0]
Domain: Modular Symbols space of dimension 5
   for Gamma_0(23) of weight ...
Codomain: Space of Boundary Modular Symbols for
   Congruence Subgroup Gamma0(23) ...
sage:
   delta=M.boundary_map().matrix().transpose()
   ; delta
[ 1  0  0  0  0]
[-1  0  0  0  0]
```

By default, Sage outputs the transpose of the matrix of a linear map. In this case, we see that the last four elements of the basis of $\mathbb{M}(\Gamma_0(23))$ form a basis of $\mathbb{S}(\Gamma_0(23))$ (which is the kernel of δ). In general, we can compute it with Sage as follows:

```
sage: S=M.cuspidal_submodule()
sage: S.basis()
((1,17), (1,19), (1,20), (1,21))
```

We obtain that the Manin symbols $(1:17)$, $(1:19)$, $(1:20)$ and $(1:21)$ form a basis of $\mathbb{S}(\Gamma_0(23))$. In general, the elements of a basis of a subspace of cuspidal modular symbols are not necessarily Manin symbols, but formal sums of Manin symbols.

Now we can compute the action of some Hecke operators on $\mathbb{S}(\Gamma_0(23))$ and use it to determine the first coefficients of the q–expansions of the elements of a basis of $S_2(\Gamma_0(23))$.

The matrices of the Hecke operators with respect to the basis of $\mathbb{S}(\Gamma_0(23))$ which we have computed can be obtained in the following way:

```
sage: S.T(2)
Hecke operator T_2 on Modular Symbols subspace
   of dimension 4 of Modular Symbols space of
   dimension 5 for Gamma_0(23) of weight 2 with
   sign 0 over Rational Field
sage: T2=S.T(2).matrix().transpose() ; T2
[ 0  0 -1 -1]
[ 1  1  2  1]
[-1 -1 -2  0]
[ 0  1  1 -1]
sage: T3=S.T(3).matrix().transpose() ; T3
[-1  0  2  2]
[-2 -3 -4 -2]
[ 2  2  3  0]
[ 0 -2 -2  1]
```

Let us check that these are in fact the desired matrices:

```
sage: e2=S.basis()[1] ; e2
(1,19)
sage: fe2=S.T(2)(e2) ; fe2
(1,19) - (1,20) + (1,21)
sage: S.coordinate_vector(fe2)
(0, 1, -1, 1)
sage: T2*S.coordinate_vector(e2)
```

```
(0, 1, -1, 1)
sage: S.coordinate_vector(S.T(3)(e2))==
   T3*S.coordinate_vector(e2)
True
sage: T2*T3-T3*T2==0
True
```

Now we compute the q–expansions of 2 linearly independent elements of $S_2(\Gamma_0(23))$ up to precision $O(q^6)$ (this is going to be a basis) using the algorithm explained in the previous section:

```
sage: R.<q>=PowerSeriesRing(QQ)
sage: f00=sum(S.T(n).matrix()[0,0]*q^n for n in
   xrange(1,6))+O(q^6) ; f00
q - q^3 - q^4 + O(q^6)
sage: f01=sum(S.T(n).matrix()[1,0]*q^n for n in
   xrange(1,6))+O(q^6) ; f01
O(q^6)
sage: f10=sum(S.T(n).matrix()[0,1]*q^n for n in
   xrange(1,6))+O(q^6) ; f10
q^2 - 2*q^3 - q^4 + 2*q^5 + O(q^6)
```

We have computed the first 5 coefficients of the q–expansions of the three elements $f_{00}, f_{01}, f_{10} \in S_2(\Gamma_0(23))$. It is clear that f_{00} and f_{10} are linearly independent and, consequently, $f_{01} = 0$. In conclusion, the complex vector space $S_2(\Gamma_0(23))$ has a basis consisting of the cusp forms f_{00} and f_{10}, whose q–expansions are

$$\widehat{(f_{00})}_\infty(q) = q - q^3 - q^4 + O(q^6) \quad \text{and} \quad \widehat{(f_{10})}_\infty(q) = q^2 - 2q^3 - q^4 + 2q^5 + O(q^6).$$

Alternatively, the command `S.q_expansion_cuspforms` returns a function `f` such that `f(i,j)` is the q–expansion of f_{ij} to some precision. (In fact, Sage uses the matrices of the Hecke operators acting on a basis of the dual space of $\mathbb{S}(\Gamma_0(23))$ and so obtains another basis.)

```
sage: f=S.q_expansion_cuspforms(6)
sage: f(0,0)
q - 2/3*q^2 + 1/3*q^3 - 1/3*q^4 - 4/3*q^5 +
   O(q^6)
sage: f(0,1)
O(q^6)
```

```
sage: f(1,0)
-1/3*q^2 + 2/3*q^3 + 1/3*q^4 - 2/3*q^5 + O(q^6)
```

Another possibility is to use the command `S.q_expansion_basis`, which returns a basis in echelon form:

```
sage: S.q_expansion_basis(6)
[
q - q^3 - q^4 + O(q^6)
q^2 - 2*q^3 - q^4 + 2*q^5 + O(q^6)
]
```

In this particular case, we can also use the structure of $S_2(\Gamma_0(23))$ to compute a different basis. Since 23 is prime, there are no oldforms (i.e., cusp forms arising from lower levels dividing 23). Moreover, the space of newforms (which is the whole $S_2(\Gamma_0(23))$) decomposes as the direct sum of two one-dimensional eigenspaces: let us check it.

```
sage: T2.charpoly().factor()
(x^2 + x - 1)^2
```

The characteristic polynomial of $[T(2)]$ is $(X^2+X-1)^2$. Its roots, $\frac{-1+\sqrt{5}}{2}$ and $\frac{-1-\sqrt{5}}{2}$, are defined in the quadratic field $K = \mathbb{Q}(\sqrt{5})$. Therefore, we can compute an eigenvector with coefficients in K for each of these two eigenvalues:

```
sage: K.<sqrt5>=NumberField(x^2-5) ; K
Number Field in sqrt5 with defining polynomial
   x^2 - 5
sage: T2ext=matrix(K,T2) ; T2ext
[ 0  0 -1 -1]
[ 1  1  2  1]
[-1 -1 -2  0]
[ 0  1  1 -1]
sage: T2ext.charpoly().factor()
(x - 1/2*sqrt5 + 1/2)^2 * (x + 1/2*sqrt5 +
   1/2)^2
sage: T2ext.eigenvectors_right()
[
(1/2*sqrt5 - 1/2, [
  (1, 0, 1/2*sqrt5 - 3/2, -sqrt5 + 2),
  (0, 1, 1/2*sqrt5 - 3/2, -1/2*sqrt5 + 3/2)
```

```
  ], 2),
(-1/2*sqrt5 - 1/2, [
  (1, 0, -1/2*sqrt5 - 3/2, sqrt5 + 2),
  (0, 1, -1/2*sqrt5 - 3/2, 1/2*sqrt5 + 3/2)
  ], 2)
]
sage: u1=T2ext.eigenvectors_right()[0][1][0] ;
   u1
(1, 0, 1/2*sqrt5 - 3/2, -sqrt5 + 2)
sage: u2=T2ext.eigenvectors_right()[1][1][0] ;
   u2
(1, 0, -1/2*sqrt5 - 3/2, sqrt5 + 2)
```

The two eigenvectors (with distinct eigenvalues) u_1 and u_2 of the matrix $[T(2)]$ correspond to eigenforms of the Hecke operator $T(2)$ acting on $S_2(\Gamma_0(23))$ (because Hecke operators are defined using the same set of matrices on $S_2(\Gamma_0(23))$ and on $\mathbb{S}(\Gamma_0(23))$). Since the Hecke operators commute, every $T(n)$ (for $n \in \mathbb{N}$) preserves the eigenspaces of $T(2)$. In particular, since the eigenspaces of $T(2)$ are one-dimensional, u_1 and u_2 correspond to eigenforms of all the Hecke operators. Moreover, since the matrices $[T(n)]$ have integer entries and the components of u_1 and of u_2 are in K, the eigenvalues of $T(n)$ are in K for all $n \in \mathbb{N}$. These eigenvalues are the Fourier coefficients of the corresponding eigenform. Indeed, since the first component of both u_1 and u_2 is 1, we can define for $i \in \{1,2\}$ the element $\widehat{u_i}$ of $\mathrm{Hom}_{\mathbb{C}}(\mathbb{T}_{\mathbb{C}},\mathbb{C})$ which maps a Hecke operator $T(n)$ to the first component of $[T(n)]u_i$ (which is the corresponding eigenvalue of $T(n)$); by proposition 4.3, the coefficients of the q–expansion of the associated cusp form are given by the images of the Hecke operators under $\widehat{u_i}$. Therefore, all the computations can be performed in the field K. (A similar argument shows that, in general, the Fourier coefficients of a normalised eigenform of all the Hecke operators are defined in a finite extension of $\mathbb{Q}$.)

Having argued that we can compute a basis of eigenforms with Fourier coefficients in K, we can use Sage to actually perform all the computations. To do so, we need to define $\mathbb{M}(\Gamma_0(23)) \otimes_{\mathbb{Z}} K$ and all the associated objects obtained by extension of scalars:

```
sage: R.<q>=PowerSeriesRing(K)
sage: Mext=ModularSymbols(G,base_ring=K) ; Mext
Modular Symbols space of dimension 5 for
```

```
    Gamma_0(23) of weight 2 with sign 0 over
    Number Field in sqrt5 with defining
    polynomial x^2 - 5
sage: Sext=Mext.cuspidal_submodule()
sage: Sext.basis()
((1,17), (1,19), (1,20), (1,21))
sage: Sext.T(2).matrix().transpose()==T2ext
True
```

Since K is a totally real field, not only are the cusp forms associated with u_1 and u_2 eigenforms, but also u_1 and u_2 are eigenvectors of the matrices $[T(n)]$ for $n \in \mathbb{N}$ (recall that the isomorphism between $\mathbb{S}(\Gamma_0(23)) \otimes_{\mathbb{Z}} \mathbb{R}$ and $S_2(\Gamma_0(23))$ is an isomorphism of real vector spaces). We check it for some cases:

```
sage: T3ext=Sext.T(3).matrix().transpose()
sage: T5ext=Sext.T(5).matrix().transpose()
sage: T3ext*u1
(-sqrt5, 0, 3/2*sqrt5 - 5/2, -2*sqrt5 + 5)
sage: T3ext*u1==-sqrt5*u1
True
sage: T3ext*u2==(T3ext*u2)[0]*u2
True
sage: T5ext*u1==(T5ext*u1)[0]*u1
True
sage: T5ext*u2==(T5ext*u2)[0]*u2
True
```

Finally, we compute the q–expansions of the associated cusp forms to precision $O(q^{11})$:

```
sage:
    f1=sum((Sext.T(n).matrix().transpose()*u1)[0]
    *q^n for n in xrange(1,11)) + O(q^11)
sage: f1
q + (1/2*sqrt5 - 1/2)*q^2 - sqrt5*q^3 +
    (-1/2*sqrt5 - 1/2)*q^4 + (sqrt5 - 1)*q^5 +
    (1/2*sqrt5 - 5/2)*q^6 + (sqrt5 + 1)*q^7 -
    sqrt5*q^8 + 2*q^9 + (-sqrt5 + 3)*q^10 +
    O(q^11)
```

```
sage:
   f2=sum((Sext.T(n).matrix().transpose()*u2)[0]
   *q^n for n in xrange(1,11)) + O(q^11)
sage: f2
q + (-1/2*sqrt5 - 1/2)*q^2 + sqrt5*q^3 +
   (1/2*sqrt5 - 1/2)*q^4 + (-sqrt5 - 1)*q^5 +
   (-1/2*sqrt5 - 5/2)*q^6 + (-sqrt5 + 1)*q^7 +
   sqrt5*q^8 + 2*q^9 + (sqrt5 + 3)*q^10 +
   O(q^11)
```

All in all, the complex vector space $S_2(\Gamma_0(23))$ has a basis consisting of the two eigenforms f_1 and f_2 of all the Hecke operators, whose q–expansions are

$$\widehat{(f_1)}_\infty(q) = q - \frac{1-\sqrt{5}}{2}q^2 - \sqrt{5}q^3 - \frac{1+\sqrt{5}}{2}q^4 - \left(1-\sqrt{5}\right)q^5 - \frac{5-\sqrt{5}}{2}q^6 + O(q^7)$$

and

$$\widehat{(f_2)}_\infty(q) = q - \frac{1+\sqrt{5}}{2}q^2 + \sqrt{5}q^3 - \frac{1-\sqrt{5}}{2}q^4 - \left(1+\sqrt{5}\right)q^5 - \frac{5+\sqrt{5}}{2}q^6 + O(q^7).$$

Observe that the Fourier coefficients of these two q–expansions are Galois conjugates (that is, they only differ in the sign of the square root of 5).

5.4 Modular symbols for $\Gamma_0(77)$

In this section, we illustrate some of the computations for $\Gamma_0(77)$. However, we focus on the results obtained using Sage [15] and do not discuss further the algorithms (in contrast with the previous section).

First, we define our spaces of modular symbols and compute a basis:

```
sage: G=Gamma0(77)
sage: M=ModularSymbols(G)
sage: S=M.cuspidal_submodule()
sage: M.basis()
((1,0),
 (1,74),
 (1,75),
 (7,1),
 (7,3),
```

```
 (7,5),
 (7,6),
 (7,8),
 (7,9),
 (7,10),
 (11,1),
 (11,2),
 (11,3),
 (11,4),
 (11,5),
 (11,6),
 (11,7))
sage: S.basis()
((1,74),
 (1,75),
 (7,1) - (11,6) + (11,7),
 (7,3) - (11,6) + (11,7),
 (7,5) - (11,6) + (11,7),
 (7,6) - (11,6) + (11,7),
 (7,8) - (11,6) + (11,7),
 (7,9) - (11,6) + (11,7),
 (7,10) - (11,6) + (11,7),
 (11,1) - (11,6),
 (11,2) - (11,6),
 (11,3) - (11,6),
 (11,4) - (11,6),
 (11,5) - (11,6))
```

Observe that, in this case, not all the elements of the basis of $\mathbb{S}(\Gamma_0(77))$ are Manin symbols (there are sums of several Manin symbols as well).

The rank of $\mathbb{S}(\Gamma_0(77))$ is 14, so the (complex) dimension of $S_2(\Gamma_0(77))$ must be 7. We can compute the q–expansions of the elements of a basis of $S_2(\Gamma_0(77))$ to precision $O(q^{13})$ as follows:

```
sage: S.q_expansion_basis(13)
[
q - 2/5*q^8 + 2/5*q^9 - 6/5*q^10 - 1/5*q^11 -
   2/5*q^12 + O(q^13),
```

```
q^2 + 3/5*q^8 - 8/5*q^9 - 6/5*q^10 - 1/5*q^11 -
   12/5*q^12 + O(q^13),
q^3 - 3/5*q^8 - 2/5*q^9 + 1/5*q^10 + 1/5*q^11 -
   3/5*q^12 + O(q^13),
q^4 + 2/5*q^8 - 7/5*q^9 - 4/5*q^10 + 1/5*q^11 -
   8/5*q^12 + O(q^13),
q^5 + 3/5*q^8 - 8/5*q^9 - 1/5*q^10 - 1/5*q^11 -
   7/5*q^12 + O(q^13),
q^6 - 1/5*q^8 - 4/5*q^9 - 3/5*q^10 + 2/5*q^11 -
   6/5*q^12 + O(q^13),
q^7 + O(q^13)
]
```

In this case, though, we cannot find a basis of eigenforms as easily as in the case of $\Gamma_0(23)$. To do so, we would have to study the oldforms arising from elements of $S_2(\Gamma_0(7))$ and of $S_2(\Gamma_0(11))$, but we have not presented the required theory. Instead, we use Sage to compute a basis of eigenforms of the new subspace of $S_2(\Gamma_0(77))$ (that is, the subspace generated by the newforms).

We define the subspace $\mathbb{S}^{\text{new}}(\Gamma_0(77))$ of $\mathbb{S}(\Gamma_0(77))$ corresponding to the new subspace $S_2^{\text{new}}(\Gamma_0(77))$ of $S_2(\Gamma_0(77))$ and study its decomposition as a direct sum of its eigenspaces:

```
sage: Sn=S.new_submodule()
sage: Sn.basis()
((1,74) + (7,9) + (7,10) - (11,3) - (11,5) +
   2*(11,7),
 (1,75) + (7,9) - (11,5) + (11,7),
 (7,1) + (7,9) + (7,10) - (11,3) - (11,5) -
    (11,6) + 3*(11,7),
 (7,3) + (7,9) - (11,5) - (11,6) + 2*(11,7),
 (7,5) + (7,10) - (11,3) - (11,6) + 2*(11,7),
 (7,6) - (7,9) - (7,10) + (11,5) - (11,7),
 (7,8) + (7,9) - (11,3) - (11,5) + 2*(11,7),
 (11,1) - (11,3),
 (11,2) - (11,5),
 (11,4) - (11,6))
sage: Sn.decomposition()
[
```

```
Modular Symbols subspace of dimension 2 of
   Modular Symbols space of dimension 17 for
   Gamma_0(77) of weight 2 with sign 0 over
   Rational Field,
Modular Symbols subspace of dimension 2 of
   Modular Symbols space of dimension 17 for
   Gamma_0(77) of weight 2 with sign 0 over
   Rational Field,
Modular Symbols subspace of dimension 2 of
   Modular Symbols space of dimension 17 for
   Gamma_0(77) of weight 2 with sign 0 over
   Rational Field,
Modular Symbols subspace of dimension 4 of
   Modular Symbols space of dimension 17 for
   Gamma_0(77) of weight 2 with sign 0 over
   Rational Field
]
```

There is a subspace of dimension 4 of $\mathbb{S}^{\text{new}}(\Gamma_0(77)) \otimes_{\mathbb{Z}} \mathbb{Q}$ which cannot be further decomposed as a rational vector space. Hence, we cannot apply the reasoning which we used to prove that $S_2(\Gamma_0(23))$ decomposes as the direct sum of one-dimensional eigenspaces in section 5.3: first, we need to find the appropriate extension of scalars.

We can determine heuristically the (finite) field extension K of $\mathbb{Q}$ such that $\mathbb{S}^{\text{new}}(\Gamma_0(77)) \otimes_{\mathbb{Z}} K$ can be decomposed as the direct sum of two-dimensional eigenspaces (each of which corresponds to the subspace of $S_2(\Gamma_0(77))$ generated by a newform which is an eigenform of all the Hecke operators). That is to say, we compute the characteristic polynomial of some Hecke operator acting on $\mathbb{S}^{\text{new}}(\Gamma_0(77))$ and determine its splitting field.

```
sage: Sn.T(2).charpoly().factor()
(x - 1)^2 * x^4 * (x^2 - 5)^2
```

Hence, we define $K = \mathbb{Q}(\sqrt{5})$ and check that $\mathbb{S}^{\text{new}}(\Gamma_0(77)) \otimes_{\mathbb{Z}} K$ decomposes as the direct sum of two-dimensional eigenspaces:

```
sage: K.<sqrt5>=NumberField(x^2-5)
sage: MK=ModularSymbols(G,base_ring=K)
sage: SK=MK.cuspidal_submodule()
```

```
sage: SnK=SK.new_subspace()
sage: SnK.decomposition()
[
Modular Symbols subspace of dimension 2 of
   Modular Symbols space of dimension 17 for
   Gamma_0(77) of weight 2 with sign 0 over
   Number Field in sqrt5 with defining
   polynomial x^2 - 5,
Modular Symbols subspace of dimension 2 of
   Modular Symbols space of dimension 17 for
   Gamma_0(77) of weight 2 with sign 0 over
   Number Field in sqrt5 with defining
   polynomial x^2 - 5,
Modular Symbols subspace of dimension 2 of
   Modular Symbols space of dimension 17 for
   Gamma_0(77) of weight 2 with sign 0 over
   Number Field in sqrt5 with defining
   polynomial x^2 - 5,
Modular Symbols subspace of dimension 2 of
   Modular Symbols space of dimension 17 for
   Gamma_0(77) of weight 2 with sign 0 over
   Number Field in sqrt5 with defining
   polynomial x^2 - 5,
Modular Symbols subspace of dimension 2 of
   Modular Symbols space of dimension 17 for
   Gamma_0(77) of weight 2 with sign 0 over
   Number Field in sqrt5 with defining
   polynomial x^2 - 5
]
```

Finally, we can compute the q–expansions of these eigenforms to precision $O(q^{11})$ as follows:

```
sage: SnK[0].q_eigenform(11,'a')
q - sqrt5*q^2 + (sqrt5 + 1)*q^3 + 3*q^4 - 2*q^5
   + (-sqrt5 - 5)*q^6 + q^7 - sqrt5*q^8 +
   (2*sqrt5 + 3)*q^9 + 2*sqrt5*q^10 + O(q^11)
sage: SnK[1].q_eigenform(11,'a')
```

```
q - 3*q^3 - 2*q^4 - q^5 - q^7 + 6*q^9 + O(q^11)
sage: SnK[2].q_eigenform(11,'a')
q + q^3 - 2*q^4 + 3*q^5 + q^7 - 2*q^9 + O(q^11)
sage: SnK[3].q_eigenform(11,'a')
q + q^2 + 2*q^3 - q^4 - 2*q^5 + 2*q^6 - q^7 -
   3*q^8 + q^9 - 2*q^10 + O(q^11)
sage: SnK[4].q_eigenform(11,'a')
q + sqrt5*q^2 + (-sqrt5 + 1)*q^3 + 3*q^4 -
   2*q^5 + (sqrt5 - 5)*q^6 + q^7 + sqrt5*q^8 +
   (-2*sqrt5 + 3)*q^9 - 2*sqrt5*q^10 + O(q^11)
```

As expected, all but 2 of these q–expansions have coefficients in $\mathbb{Q}$.

Alternatively, we could have computed these q–expansions directly with Sage without previously knowing the field of definition of the Fourier coefficients:

```
sage: Sn[0].q_eigenform(11,'a')
q - 3*q^3 - 2*q^4 - q^5 - q^7 + 6*q^9 + O(q^11)
sage: Sn[1].q_eigenform(11,'a')
q + q^3 - 2*q^4 + 3*q^5 + q^7 - 2*q^9 + O(q^11)
sage: Sn[2].q_eigenform(11,'a')
q + q^2 + 2*q^3 - q^4 - 2*q^5 + 2*q^6 - q^7 -
   3*q^8 + q^9 - 2*q^10 + O(q^11)
sage: Sn[3].q_eigenform(11,'a')
q + a*q^2 + (-a + 1)*q^3 + 3*q^4 - 2*q^5 + (a -
   5)*q^6 + q^7 + a*q^8 + (-2*a + 3)*q^9 -
   2*a*q^10 + O(q^11)
sage: f=Sn[3].q_eigenform(11,'alpha') ; f
q + alpha*q^2 + (-alpha + 1)*q^3 + 3*q^4 -
   2*q^5 + (alpha - 5)*q^6 + q^7 + alpha*q^8 +
   (-2*alpha + 3)*q^9 - 2*alpha*q^10 + O(q^11)
sage: f.base_ring()
Number Field in alpha with defining polynomial
   x^2 - 5
```

These last lines indicate that there are two eigenforms whose q–expansions are of the form

$$q + \alpha q^2 + (1-\alpha)q^3 + 3q^4 - 2q^5 + (\alpha - 5)q^6 + q^7 + \alpha q^8 + (3 - 2\alpha)q^9 - 2\alpha q^{10} + O(q^{11})$$

for the roots α of the polynomial $X^2 - 5$ (that is, $\alpha = \pm\sqrt{5}$, coinciding with our previous computations).

Bibliography

[1] Cremona, J. E. *Algorithms for modular elliptic curves*. 2nd ed. New York, NY, USA: Cambridge University Press, 1997. 376 pp. URL: http://homepages.warwick.ac.uk/~masgaj/book/fulltext/index.html (visited on 06/09/2015).

[2] Diamond, F. and Shurman, J. *A first course in modular forms*. Graduate texts in mathematics 228. New York, NY, USA: Springer Science+Business Media, 2005. 436 pp.

[3] Koblitz, N. *Introduction to elliptic curves and modular forms*. 2nd ed. Graduate texts in mathematics 97. New York, NY, USA: Springer–Verlag, 1993. 248 pp.

[4] Lang, S. *Introduction to modular forms*. Grundlehren der mathematischen Wissenschaften 222. Berlin, Germany: Springer–Verlag, 1995. 261 pp. Corrected reprint of the 1976 original.

[5] Manin, Y. I. "Parabolic points and zeta–functions of modular curves". In: *Mathematics of the USSR-Izvestiya* 6.1 (1972), pp. 19–64.

[6] Merel, L. "Universal Fourier expansions of modular forms". In: *On Artin's conjecture for odd 2–dimensional representations*. Lecture notes in mathematics 1585. Berlin, Germany: Springer–Verlag, 1994, pp. 59–94. URL: http://webusers.imj-prg.fr/~loic.merel/Recherche.html (visited on 06/09/2015).

[7] Miatello, R. J. *Formas modulares y operadores de Hecke*. Course notes. 2015. 19 pp. URL: http://webusers.imj-prg.fr/~harald.helfgott//agraweb/AGRAIIMiatello.pdf (visited on 06/09/2015).

[8] Milne, J. S. *Modular functions and modular forms (v1.30)*. Course notes. 2012. 138 pp. URL: http://www.jmilne.org/math/CourseNotes/mf.html (visited on 06/09/2015).

[9] Miyake, T. *Modular forms*. Springer monographs in mathematics. Berlin, Germany: Springer–Verlag, 2006. 335 pp. Corrected reprint of the 1989 original.

[10] Reyssat, E. *Quelques aspects des surfaces de Riemann*. Progress in mathematics 77. Boston, MA, USA: Birkhäuser, 1989. 166 pp.

[11] Serre, J.-P. *A course in arithmetic*. Graduate texts in mathematics 7. New York, NY, USA: Springer–Verlag, 1973. 115 pp.

[12] Shimura, G. *Introduction to the arithmetic theory of automorphic functions*. Kanô memorial lectures 1. Princeton, NJ, USA: Princeton University Press, 1994. 271 pp. Reprint of the 1971 original.

[13] Stein, W. A. "An introduction to computing modular forms using modular symbols". In: *Algorithmic number theory. Lattices, number fields, curves and cryptography*. Mathematical Sciences Research Institute publications 44. New York, NY, USA: Cambridge University Press, 2008, pp. 641–652. URL: http://library.msri.org/books/Book44/contents.html (visited on 10/09/2015).

[14] Stein, W. A. *Modular forms, a computational approach*. Graduate studies in mathematics 79. Providence, RI, USA: American Mathematical Society, 2007. 268 pp. URL: http://wstein.org/books/modform/ (visited on 26/08/2015).

[15] Stein, W. A. et al. *Sage mathematics software*. Version 7.1. 2016. URL: http://www.sagemath.org (visited on 28/03/2016).

Indices

General index

Index of symbols

Symbols